

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
Art Unit: _____ Phone Number 30 _____ Serial Number: _____
Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: <u>1664 Port</u>	NA Sequence (#) <u>1</u>	STN _____
Searcher Phone #: <u>301-3537</u>	AA Sequence (#) <u>1</u>	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: <u>4/11</u>	Bibliographic _____	Dr.Link _____
Date Completed: <u>4/13</u>	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: <u>15</u>	Fulltext _____	Sequence Systems <u>gs</u>
Clerical Prep Time: _____	Patent Family _____	WWW/Internet _____
Online Time: <u>10</u>	Other _____	Other (specify) _____

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2001, 16:22:00 ; Search time 18.58 Seconds

(without alignments)
1546.081 Million cell updates/sec

Title: US-09-464-039-7

Perfect score: 2139
Sequence: 1 MLPNTGLACTVFTTCASR.....GNMALAIKLEKMNQMARL 418

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1034	48.3	938	2 T34105	hypothetical prote
2	891.5	41.7	293	2 T19954	hypothetical prote
3	708	33.1	274	2 G83284	probable short-cha
4	665	31.1	282	2 A70597	hypothetical prote
5	286	13.4	436	2 T16638	hypothetical prote
6	267	12.5	248	2 T44932	3-oxoacyl-(acyl)-ca
7	256	12.0	238	1 D69930	probable 3-oxoacyl
8	244	11.4	240	1 H75014	3-oxoacyl-(acyl)-ca
9	241	11.3	246	2 H72219	glucose 1-dehydrog
10	237	11.1	258	2 G69755	estradiol 17beta-d
11	236	11.0	736	2 S59136	3-oxoacyl-(acyl)-ca
12	235	11.0	246	2 C83961	probable short-cha
13	235	11.0	266	2 F83127	hypothetical prote
14	233.5	10.9	255	2 D70635	hypothetical prote
15	232	10.8	278	2 T10877	Y41A protein - Rhl
16	231	10.8	894	2 S54786	multifunctional be
17	227	10.6	679	2 T36426	probable oxidoredu
18	224.5	10.5	301	2 H70890	hypothetical prote
19	224	10.5	285	1 A65017	probable 3-oxoacyl
20	223	10.4	275	2 T14021	probable sterol de
21	222.5	10.4	271	1 S34678	short-chain alcoh
22	222.5	10.4	320	2 S22450	3-oxoacyl-(acyl)-ca
23	222	10.4	253	2 E83207	probable short-cha
24	221	10.3	248	2 D70707	hypothetical prote
25	218.5	10.2	246	2 A68621	3-oxoacyl-(acyl)-ca
26	218.5	10.2	251	2 G72389	oxidoreductase, sh
27	217	10.1	256	2 A70637	hypothetical prote
28	217	10.1	261	2 D69629	glucose 1-dehydrog
29	216.5	10.1	255	2 T44991	oxidoreductase [m

30	214.5	10.0	246	2 H84136	3-oxoacyl-(acyl)-ca
31	214.5	10.0	284	2 T28018	hypothetical prote
32	212.5	9.9	735	2 S74209	multifunctional be
33	211.5	9.9	241	1 E47069	(R,R)-butanediol d
34	211	9.9	279	2 T32196	hypothetical prote
35	211	9.9	900	1 S25322	bifunctional beta-
36	210.5	9.8	285	2 T50932	short-chain dehydr
37	208	9.7	265	2 T36396	probable short cha
38	206.5	9.7	286	2 D69768	alcohol dehydrogen
39	206	9.6	250	1 S47054	probable dehydroge
40	206	9.6	256	2 E83396	3-hydroxybutyrate
41	206	9.6	592	2 G70786	probable epnd prot
42	205.5	9.6	255	2 G83644	2,5-dichloro-2,5-c
43	205.5	9.6	319	2 S22416	3-oxoacyl-(acyl)-ca
44	205	9.6	250	2 C83622	probable short-cha
45	204.5	9.6	253	1 S17711	probable dehydroge

ALIGNMENTS

RESULT	1	Query Match	Score	1034;	DB 2;	Length	938;
T34105	hypothetical protein C17G10.8 - Caenorhabditis elegans	Best Local Similarity	54.6%	Pred. No. 5e-68;			
C:Species: Caenorhabditis elegans		Matches	230;	Conservative	52;	Mismatches	111;
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999						Indels	28;
C:Accession: T34105						Gaps	8;
R:Johnson, D.							
A:Description: The sequence of C. elegans cosmid C17G10.							
A:Reference number: Z21476							
A:Accession: T34105							
A:Status: preliminary; translated from GB/EMBL/DBJ							
A:Molecule type: DNA							
A:Residues: 1-938 <JOB>							
A:Cross-references: EMBL:U28739; PIDN:AAB93456.1; GSPDB:GN00020; CESP:C17G10.8							
A:Experimental source: strain Bristol NZ; clone C17G10							
C:Genetics:							
A:Gene: CESP:C17G10.8							
A:Map position: 2							
A:Introns: 25/1; 67/2; 158/3; 214/3; 329/3; 386/2; 465/3; 510/1; 525/2; 582/3; 709/1;							
OY	7	RLAGCTVFTTGASRGIGKATLAKADGANIVTAATQAPHPKLTGTIYTAEEIEAVG	66				
DB	526	KFVGRVLTITGASRGIGKATLAKADGANIVTAATQAPHPKLTGTIYTAEEIEAVG	585				
OY	67	KALPCIVDVRDEQOISAAVEAKIKRFGIDILVNNASISLNTLTPPKRLDLMNVNT	126				
DB	586	KALPCIVDVRDEQOISAAVEAKIKRFGIDILVNNASISLNTLTPPKRLDLMNVNT	645				
OY	127	KGYTLASKACIPFLPKSKSKAHIIINISPLTNVWPKOCATYIATKGSMTVLGAEE	186				
DB	646	KGYTLASKACIPFLPKSKSKAHIIINISPLTNVWPKOCATYIATKGSMTVLGAEE	705				
OY	187	KGE-IVNMLMPKTAIHTTAMDLGGPESQCRKVDIADAAYSIFOK-PRSFQNFVI	244				
DB	706	RPGIVNMLMPKTAIHTTAMDLGGPESQCRKVDIADAAYSIFOK-PRSFQNFVI	765				
OY	245	DENILKEEGIENPDVYAIKRGHPLODFLEDEYPAVSKVSTGAVPEKKEKLOLPK	304				
DB	766	DEIILAEVGTDRYACVPDAPLMPDFI---PAGYDHKKESSGA-----QIGKK	813				
OY	305	PR---SGAVEE---PFRIYKDSLDDVYKATQATILFELSGDGGT-----WFLDKS	352				
DB	814	NKTHEAGVVEEIKQIFTSKRLNLDIVKTCGFVEYFLL--KDPPTKERRITLTLKNG	871				
OY	353	GGWVGGEPSDDQADVVMSTTDDFVFMFGSKLPTNAFNGSKRLIKGNMALAIKLEKLN	412				

Db 872 EGALTDKKASGADYKFTLAPENFAPLFTGKRLPTTALMTKLOISGDMGAMKLESILR 931
 QY 413 Q 413
 Db 932 K 932

RESULT 2

T19954

hypothetical protein C45B11.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000

C:Accession: T19954

R:McMurray, A.

submitted to the EMBL Data Library, June 1996

A:Reference number: Z19202

A:Accession: T19954

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-293 <WLL>

A:Cross-references: EMBL:Z74029; PIDN:CAA98431.1; GSPDB:GN00023; CESP:C45B11.3

A:Experimental source: clone C45B11

C:Genetics:

A:Gene: CESP:C45B11.3

A:Map position: 5

A:Insertions: 9/2; 66/3

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 41.78; Score 891.5; DB 2; Length 293;

Best Local Similarity 61.78; Pred. No. 3e-58;

Matches 185; Conservative 29; Mismatches 73; Indels 13; Gaps 3;

QY 2 LPNTRLACGVTFTGASRGICAKAIAKADGANIVIAKTAOPHPKLLGTYTAAEEI 61

Db 5 IKNGFGAPKVFITGASRGIGKELALKADGANIVIAKTAOPHPKLLGTYTAAEEI 64

QY 62 EAVGKALPCIVDVRDEQISAAVEKAIKKGIDILVNNASISLTNTLDTPPKRLDLM 121

Db 65 EKAGGHALPCVVDVRDEAAVAAVAVKKFGIDILVNNASISLTNTEDTDMKRYDLM 124

QY 122 MNVTGRLYASKACIPYLKSKVAHLNLTSPPLNLPVFKOHCAVTIAKYGSMYVLG 181

Db 125 HSITRGTYLTCTCLPLKGNPHVNLSPPLDMKMGPGHVGTMAFGSMCYLG 184

QY 182 MAEEFKG-EIAVNALMPKTAIHTAAMDMLGGPGIESQCRKVDIADAAYSIFQK-PRKFT 239

Db 185 HHEFRPFGYIAVNALMPPLTAIWTSAEMFLSGDSAGNRKASIMADSAVATLSDSKKFT 244

QY 240 GNFYIDENILKEEGIEENFDVYAIKPGHPLODFLDETPPEAVSKKVESTGAVPEKBEKL 299

Db 245 GNFLIDELLOTGTITNFEQYECVGSLLIPDFEVPKSFSA-----PRFGSKL 293

RESULT 3

G83284

Probable short-chain dehydrogenase PA2892 [Imported] - Pseudomonas aeruginosa (strain PA

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: G83284

R:Stover, C.K.; Pham, X.O.; Errin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

Nature 406: 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: AB2950; MUID:20437337

A:Accession: G83284

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-274 <STO>

A:Cross-references: GB:AE004715; GB:AE004091; NID:9948977; PIDN:AAG06280.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA2892

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 33.18; Score 708; DB 2; Length 274;

Best Local Similarity 54.08; Pred. No. 8.5e-45;

Matches 147; Conservative 44; Mismatches 77; Indels 4; Gaps 4;

QY 8 LACCTVEITGASRGICAKAIAKADGANIVIAKTAOPHPKLLGTYTAAEEIAVAGK 67

Db 3 LHGKTLFTGASRGIGRIARARADGANIVIAKSAEPHKLEGTIFSVAAEVAAGCO 62

QY 68 ALPCIVDVRDEQISAAVEKAIKKGIDILVNNASISLTNTLDTPPKRLDLMNNVTR 127

Db 63 ALPLQIDVRDEQAAVAAAMARAERFGIDALVNNAGAIRLVGVEKLEPKRFDLMYQINTR 122

QY 128 GTYIASKACIYLLKSKVAHLNLTSPPLNLPVFKOHCAVTIAKYGSMYVLGMAEEF- 186

Db 123 AVLVCQAALPRLRRSANGHTLSLSPPLNLAGRFADHGPYTVKYGSMYLTLCMHBEFG 182

QY 187 KGEIYVNALMPKTAIHTAAMDMLGGPGIESQCRKVDIADAAYSIF-OKRPSFTGNFVI 244

Db 183 KYAISVNALMPKTAIHTAATIEFELGSRDAFRARTPAIDMAAAILTSBGRSLSGRLV 242

QY 245 DENILKEEGIEENFDVYAIKPGHPLODFELD 275

Db 243 DEELLRERGQSGFEQRYRDEPGALVDPDLFD 274

RESULT 4

A70597

hypothetical protein Rv3224 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: A70597

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garrier, T.; Churcher, C.; Harris, D.; Gordon

; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd,

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393: 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70597; MUID:98295987

A:Accession: A70597

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-282 <COL>

A:Cross-references: GB:Z95120; GB:AL123456; NID:g3261739; PIDN:CA808313.1; PID:g20726

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: Rv3224

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 31.18; Score 665; DB 2; Length 282;

Best Local Similarity 48.98; Pred. No. 1.3e-41;

Matches 133; Conservative 57; Mismatches 78; Indels 4; Gaps 4;

QY 8 LACCTVEITGASRGICAKAIAKADGANIVIAKTAOPHPKLLGTYTAAEEIAVAGK 67

Db 3 LKNGTMFTGASRGIGRIARARADGANIVIAKSAEPHKLEGTIFSVAAEVAAGCO 62

QY 68 ALPCIVDVRDEQISAAVEKAIKKGIDILVNNASISLTNTLDTPPKRLDLMNNVTR 127

Db 63 ALPLVIGIRPDAAVASAVATVEQFGIDICVNNASAINSGSITIEVPMKRDILNNGIOVR 122

QY 128 GTYIASKACIYLLKSKVAHLNLTSPPLNLPVFKOHCAVTIAKYGSMYVLGMAEEF 187

Db 123 GTYAVSQACIPHMKGREPHILITSPPLLEKKRLP-TAYMAKYGKWTLCALCIAEEMR 181

QY 188 GE-IAYNALMPKTAIHTAAMDMLGGPGIESQCRKVDIADAAYSIFQKPKS-FTGNFVI 244

Db 182 ADGIASNTLMRTWVATAAAGVNLGGDEAAMARSKPEVYADAAVIVYNNKPATEYTGKTL 241

OY 245 DENILKEGIENFDYAIKPGHPLDPFLE 276
 DB 242 CEDVLVESGVTLSDVDCVPGATLGVLDWEVD 273

RESULT 5
 716638
 hypothetical protein M03A8.1 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T16638
 R:Du, Z.
 submitted to the EMBL Data Library, November 1995

A:Description: The sequence of *C. elegans* cosmid M03A8.
 A:Reference number: Z18551
 A:Accession: T16638
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-436 <DUZ>
 A:Cross-references: EMBL:U41544; NID:g1109886; PID:g1109889; PIDN:AA83184.1; CESP:M03A8
 C:Genetics:
 A:Gene: CESP:M03A8.1
 A:Introns: 92/1; 115/1; 205/3; 361/1; 405/3

Query Match 13.4%; Score 286; DB 2; Length 436;
 Best Local Similarity 25.8%; Pred. No. 1.8e-13;
 Matches 120; Conservative 73; Mismatches 176; Indels 96; Gaps 17;

OY 7 RLACGTVFTGASRGIGKAIKALKAADGANIYAAKTAOPHRLCTIYTA-----EE 60
 DB 4 RPDGKVAIVTGAGGGLGKTYALELAKRGCKVYVNDLGDRH---CTSSSSMAKRVGE 59
 OY 61 IEAVGKALPCIVDVRDEQOISAAYEKAIRKGGIDILVNNASISLTNTLPTRKLDL 120
 DB 60 IKSAGGQAAVANDSVFEGDKI---VKTAIDNCRIDIVINNGLRDSVFLKMTLMDL 116
 OY 121 MNMVTNRGYLAKACIPYLKKSKVAHILNISPLNLNPFVFKHOCAYITAKYGMVYL 180
 DB 117 IKRVHKGKAVNTKAMPFMRDQKGRIVYVSSNAGVHNFQGAN--YAAKSAALIGLSN 174
 OY 181 GNAEE-FKGEIIVNMLP-----KTAIHTAAMDML----- 209
 DB 175 SLAOGAKNIILANTLVPAAGSRLEFETVPMONLVDALPKDQVYVPLVYVWHDSFEESGKV 234
 OY 210 ---GGG---ISQCKRVIIIDAAVSTIQKPKSFNGNVIDENT--LKEEGLENFDVA 261
 DB 235 FEAGAGWYGTIOYKSKGVISHASADDIKAMNSTITNNNGAEYIGTITEQSARLVSI-- 292
 OY 262 IKRPHLPDPFLDEY-----PEAVSKRYESTGAVPEFKEEKLQLOPKRSGAVETFR 316
 DB 293 -----LKEHSASSSSSSSSGSSGSGAPPS-----NIRSSAL---FDE 327
 OY 317 VKDSISDD--VVKATQATVYLFELSGEDG---GTWFLDLKSKGNVYGE--PSDAQDV 368
 DB 328 MADGVKADPTAVTKLSIYLIIT--DGKNELGKFTLDRKSSPSYVLDVKNKEKAMNT 385
 OY 369 MGMTDDPEYKMFSGKIKPTMAFMSGKLKIKGNMATAIKLEKIMNO 413
 DB 386 VIVADSDVFVDAIGAKLMAQKAFMSGKLKVKGNMVLQKLTQVLEK 430

RESULT 6
 T44932
 3-oxoacyl-[acyl-carrier-protein] reductase (EC 1.1.1.100) [imported] - *Agrobacterium tum*
 C:Species: *Agrobacterium tumefaciens*
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
 C:Accession: T44932
 R:Kim, K.S.; Farrand, S.K.
 J. Bacteriol. 178, 3275-3284, 1996
 A:Title: T1 plasmid-encoded genes responsible for catabolism of the crown gall opine man
 by the plant tumor.
 A:Reference number: Z22872; MUID:96236046

A:Accession: T44932
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-248 <KIM>
 A:Cross-references: EMBL:U19620; NID:g797330; PIDN:AA807783.1; PID:g797334
 A:Experimental source: strain 15955
 C:Genetics:
 A:Gene: mocc
 A:Genome: plasmid pT115955
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
 C:Keywords: oxidoreductase

Query Match 12.5%; Score 267; DB 2; Length 248;
 Best Local Similarity 34.6%; Pred. No. 2e-12;
 Matches 71; Conservative 41; Mismatches 71; Indels 22; Gaps 8;

OY 7 RLACGTVFTGASRGIGKAIKALKAADGANIYAAKTAOPHRLCTIYTAEEIYAVGC 66
 DB 2 KIQKAVITGAGRGIGRTALELKEGCVYLAIE-----LNEVAVAEYVAIS 54
 OY 67 KALPCIVDVRDEQOISAAYEKAIRKGGIDILVNNASISLTNTL-DYPTKRLDMMATN 125
 DB 55 EALALRTDVQHKSEVDALAKAFAERGAVIDIIVNNA-GVAIHNTIPNKEADMDWMMAIN 113
 OY 126 TRGTVAASKACIPYLKKSKVAHILNISPLNLNPFVFKHOCAYITAKYGMVYLGMAE- 184
 DB 114 LKGTFLCTOAFQHMCDRRHGHINVS--RAKVASAFGAIAMSKFGM---LGFOT 167
 OY 185 -EFKG-EIAVNMLMPKTAIHTAAMD 207
 DB 168 TDQEGIEFGVKA---TAVCPGAVD 188

RESULT 7
 D69930
 probable 3-oxoacyl-[acyl-carrier-protein] reductase (EC 1.1.1.100) yoxD - *Bacillus su*
 C:Species: *Bacillus subtilis*
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
 C:Accession: D69930; S01270
 R:Kunst, F.; Ogatawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.;
 A.; Ehrlich, S.D.; Emerson, P.T.; Entlian, K.D.; Errington, J.; Fabret, C.; Ferrari,
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funai, S.; Galizzi, A.; Gal
 iech, J.; Harwood, C.R.; Hensut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
 koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Roeha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scani
 A:Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
 aueuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
 T.; Whitors, P.; Wipet, A.; Yamamoto, H.; Yamane, K.; Yamano, K.; Yata, K.; Yoshida
 A:Authors: Yoshikawa, H.F.; Zumsel, E.; Yoshikawa, K.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*
 A:Reference number: A69580; MUID:98044033
 A:Accession: D69930
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-238 <KUN>
 A:Cross-references: GB:Z99114; GB:AL009126; NID:g2634230; PIDN:CAB13743.1; PID:g26342

A:Experimental source: strain 168
 A:Carriagan, C.M.; Haarsma, J.A.; Smith, M.T.; Wake, R.G.
 Nucleic Acids Res. 15, 8501-8509, 1987
 A:Title: Sequence features of the replication terminus of the *Bacillus subtilis* chrom
 A:Reference number: S01270; MUID:88040469
 A:Accession: S01270
 A:Molecule type: DNA
 A:Residues: 62-238 <CAR>
 A:Cross-references: EMBL:X06168; NID:g40205; PIDN:CAA29533.1; PID:g809662
 C:Genetics:
 A:Gene: yoxD
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

OY 122 MNVNRGYYLAKSKKIPPLKSKVAHIIINISPLPLNLNVPWFKQCAVYIAKNGSMYVLG 181
 Db 110 IDTNLAKYGFHCSKAVTRPMKQKREGRIINSSVVG-----ATGAGCAANYAA 157
 OY 182 MAEEFKGEIAVNAIMPRTAHTAAMDMLGPGIESQCKRVDIADAAVSIFQKPKSFETGN 241
 Db 158 KA-----cv 161
 OY 242 EVIDENILKEGIEINFDDYAIKPGHPLOPDFLDEIPFAVSKKVESTGAVE-FKEEKLQ 300
 Db 162 IGLTKTLARELANRNTITNAVAPG-----FIE-----TDMTGELPEDYKQAMLG 205
 OY 301 LQPKRSGAVEETFRIVKDSLDDVVKAT-QAIVL 334
 Db 206 QIPLARLQGPPEVAKAVFSLASDASVLTGOTIHV 240

 RESULT 13
 F83127
 Probable short-chain dehydrogenase PA4148 [imported] - Pseudomonas aeruginosa (strain PA
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: F83127
 R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Llm
 a, Lory, S.; Olson, M.V.
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A:Reference number: A82950; MUID:20437337
 A:Accession: F83127
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-266 <STO>
 A:Cross-references: GB:AE004831; GB:AE004091; NID:9950347; PIDN:AG07535.1; GSPDB:GN001
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA4148
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

 Query Match 11.0% Score 235; DB 2; Length 266;
 Best Local Similarity 32.0% Pred. No. 5e-10;
 Matches 65; Conservative 32; Mismatches 72; Indels 34; Gaps 6;

 OY 8 LAGCTVFITGASRGIGKAIALKAADGANIYAKTAQPHPKLLGTITYTAAE---IEA 63
 Db 11 LSSRALVLTGAGRGIRGRIALALRAGADVADDPG-----VAEEFTAAAIRS 59

 OY 64 VGGKALPCIVDRDEQOISAAVEKAIRKFFGIDILVNNASISLTNLTDPTRKRLDLMN 123
 Db 60 LGRRLTAGVGVSDSDSVRAVWERVATEFGRILDVAVNNAGYISIRKVAELSLADMDRYMN 119

 OY 124 VNTRGTYLASKACIPCELEKSKVAHLNIS-----PPLNLNVPWFQHCAYITAKGM 175
 Db 120 VNARKVFLCCQAEPLDLMQAQRKGRIVNLSTAGKVGLPDL-----AHYCAKFAVYIGF 172

 OY 176 SMYVLGMAEEF-KGEIAVNALMP 197
 Db 173 SN---ALAKVYARQGVYVNALCP 192

 RESULT 14
 D70635
 Hypothetical protein RV1928c - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: D70635
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
 Rajandram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: D70635
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-255 <COL>
A:Cross-references: GB:Z4498; GB:AL123456; NID:g2182489; PIDN:CAB06498.1; PID:g1806252
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv1928c
C:Family: rhlYcd dehydrogenase; short-chain alcohol dehydrogenase homology
F:12-194/Domain: short-chain alcohol dehydrogenase homology <SADH>
Query Match 10.9%; Score 233.5; DB 2; Length 255;
Best Local Similarity 32.2%; Pred. No. 6.1e-10;
Matches 65; Conservative 29; Mismatches 79; Indels 29; Gaps 5;
Oy 8 LAGCTVFTGASRGIGKALILKAADGANTIVIA-----AKTAQPHPKLGTITYAAEEIEAVGK 67
Db 9 LHGRKALITGASTGCKHYVALAYEAGVAIARHDLAEKL-----ADEITGSGK 61
Oy 68 ALPCIVDRDQOISAAVEKAIRKFGIDIIYNNASATSLNTDTPRKRLDLMNNVTR 127
Db 62 VVPCCDVSOHQOYVSMDOYTAELGIDIVACNNGITTVPMIDMPLEEQRLQNTVT 121
Oy 128 GTYLAKACIPYLK-----SKVAHILNISPLNLPWFVKHCAYTTAKYGS 176
Db 122 GVFELTAQAARAAKAMVKGOGVLIINFSMSGHIN-----POVSHYCA--SKAAVI 171
Oy 177 MYVIGMAEEF-KGEIYVNALMP 197
Db 172 HLTAKAAVELAPHKIRNVSP 193
RESULT 15
T10877
A:Title: Rhizobium sp. (strain NGR234) plasmid pNGR234a
A:Species: Rhizobium sp.
A:Variety: strain NGR234
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jan-2000
C:Accession: T10877
R:Feilberg, C.; Feilley, R.; Bairoch, A.; Broughton, W.J.; Rosenthal, A.; Perret, X.
Nature 387, 394-401, 1997
A:Title: Molecular basis of symbiosis between Rhizobium and legumes.
A:Reference number: Z14734; MUID:97305956
A:Accession: T10877
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-278 <FRED>
A:Cross-references: EMBL:AE000082; NID:g2182489; PID:g2182500
C:Genetics:
A:Gene: y41A
A:Genome: plasmid pNGR234a
R:Superfamily: rhlYcd dehydrogenase; short-chain alcohol dehydrogenase homology
F:7-190/Domain: short-chain alcohol dehydrogenase homology <SAD>
Query Match 10.8%; Score 232; DB 2; Length 278;
Best Local Similarity 29.5%; Pred. No. 8.8e-10;
Matches 80; Conservative 34; Mismatches 107; Indels 50; Gaps 8;
Oy 7 RLACGTVFTGASRGIGKALILKAADGANTIVIA-----AKTAQPHPKLGTITYAA 58
Db 3 RFEKGVAVVTTAGAGISGACALAIAREGGRVVAVDLDSAAIACIAQ----- 49
Oy 59 EEIEAVGKALPCIVDRDEQOISAAVEKAIRKFGIDIIYNNASATSLNTDTP 114
Db 50 --IAAEAGNALAMADIDIAQAVAALEFETABRHFGVDLLVNNASAMHLTPRDAIIDL 107
Oy 115 TKRLDLMNNVTRGTYLAKACIPYLKSKVAHILNISPLNLPWFVKHCAYTTAKY 174
Db 108 LAVVDQMTATNLKRGTLCCRAIPMTAIRGGAIVNNSSCGSL--GDTAOTSTAVSKAA 165

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM proteinh - protein search, using sw model

Run on: June 11, 2001, 16:23:30 ; Search time 11.77 Seconds

(Without alignments)
1216.550 Million cell updates/sec

Title: US-09-464-039-7

Sequence: 1 MLPTMTGRAGCTVFITGASR.....GNMALAIKLEKLMQNMARL 418

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	256	12.0	238	1 YOXD_BACSU	P14802 bacillus su
2	241	11.3	246	1 FABG_THEMA	O9x248 thermotoga
3	237	11.1	258	1 DHG2_BACSU	P80869 bacillus su
4	236	11.0	736	1 DHG4_HUMAN	P51659 homo sapien
5	232	10.8	278	1 Y4YA_RHISN	P55541 rhizobium s
6	231	10.8	894	1 FOX2_NEUCR	O01373 neurospora
7	224	10.5	285	1 UCPA_ECOLI	P37440 escherichia
8	223	10.4	275	1 YCPI_BRAJA	O45219 bradyrhizob
9	222.5	10.4	271	1 SDRI_PICAB	O08632 picea abies
10	222.5	10.4	320	1 FABG_CUPLA	P28643 cuphea lanc
11	218.5	10.2	246	1 FABG_BACSU	P51831 bacillus su
12	218.5	10.2	251	1 Y325_THEMA	O9w990 thermotoga
13	217	10.1	261	1 DHG_BACSU	P12310 bacillus su
14	211.5	9.9	241	1 BUOC_KLETE	O04520 klebsiella
15	211	9.9	900	1 FOX2_YEAST	O02207 saccharomyc
16	210.5	9.8	256	1 BUOC_KLEPN	O48436 klebsiella
17	210.5	9.8	735	1 DHG4_RAT	P97852 rattus norv
18	210	9.8	258	1 BDHA_RHIME	O86034 rhizobium m
19	206.5	9.7	285	1 GS33_BACSU	P80873 bacillus su
20	206	9.6	592	1 EPHD_MYCTU	O10402 mycobacteri
21	205.5	9.6	319	1 FABG_ARCTH	P33207 arabidopsis
22	204.5	9.6	253	1 KDUD_ERMCH	O05528 erwinia chr
23	203	9.5	255	1 HDHA_ECOLI	P25529 escherichia
24	201.5	9.4	261	1 DHG1_BACME	P39462 bacillus me
25	201.5	9.4	261	1 DHG4_BACME	P40288 bacillus me
26	200.5	9.4	735	1 DHG4_MOUSE	P51660 mus musculus
27	200	9.4	257	1 YXJF_BACSU	P42317 bacillus su
28	199.5	9.3	247	1 FAGI_SYNY3	P73574 synecocyst
29	198.5	9.3	262	1 DHGB_BACME	P07999 bacillus me
30	197.5	9.2	906	1 FOX2_CANTR	P22414 canada tiro
31	197	9.2	287	1 HETN_ANASP	P37694 anabaena sp
32	195.5	9.1	253	1 KDUD_ECOLI	P37699 escherichia
33	194	9.1	261	1 DHG2_BACME	P39483 bacillus me

34	193	9.0	250	1 LINX_PSEPA	P50198 pseudomonas
35	193	9.0	270	1 DHMA_FLASI	P22441 flavobacter
36	191.5	9.0	262	1 VER1_ASPEA	P50161 aspergillus
37	189.5	8.9	264	1 STCO_EMENTI	O00791 emericella
38	189	8.8	261	1 DHG4_BACME	P39485 bacillus me
39	189	8.8	261	1 DHG4_BACME	P10528 bacillus me
40	188.5	8.8	262	1 YXBG_BACSU	P46331 bacillus su
41	187	8.7	261	1 DHG3_BACME	P39484 bacillus me
42	181	8.5	256	1 Y019_THEMA	O56318 thermotoga
43	180.5	8.4	258	1 BDHA_ALCEU	O9x6u2 alcaligenes
44	179.5	8.4	275	1 BNZE_PSEPU	P08088 pseudomonas
45	179.5	8.4	275	1 TODD_PSEPU	P13859 pseudomonas

ALIGNMENTS

```

RESULT 1
YOXD_BACSU STANDARD; PRT; 238 AA.
ID YOXD_BACSU
AC P14802;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL OXIDOREDUCTASE IN RTP-PELB INTERGENIC REGION (EC 1.-.-.-)
DE (ORF238).
GN YOXD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=91192601; PubMed=1849493;
RA Ahn K.S., Wake R.G.;
RT "Variations and coding features of the sequence spanning the
RT replication terminus of Bacillus subtilis 168 and W23 chromosomes.";
RL Gene 98:107-112(1991).
RN [2]
RP SEQUENCE OF 62-238 FROM N.A.
RC STRAIN=168;
RX MEDLINE=88040469; PubMed=3118336;
RA Carrigan C.M., Haarsma J.A., Smith M.T., Wake R.G.;
RT "sequence features of the replication terminus of the Bacillus
RT subtilis chromosome.";
RL Nucleic Acids Res. 15:8501-8509(1987).
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X06168; CAA29533.1; -
CC EMBL; 299114; CAB13743.1; -
CC PIR; S01270; S01270.
CC HSSP; P19992; 2HSD.
CC Subtilisin; Bg11048; YOXD.
CC InterPro; IPR002196; -.
CC Pfam; PF00106; adh_short; 1.
CC PRINTS; PF00080; SDRFAMILY.
CC PROSITE; PS00061; ADH_SHORT; 1.
CC Hypothetical protein: Oxidoreductase.
FT ND_BIND 10 34 NAD OR NADP (BY SIMILARITY).
FT ACT_SITE 135 135 BY SIMILARITY.
SQ SEQUENCE 238 AA; 25299 MW; 9CC7ABD1204DF248 CRC64;

```



```

Db 174 LANSIAEGRKSNHCNTIAPNAGSMQTVMPEDELVEALKPEYVAPLVLMCHESCEEN 233
OY 191 -----AVNALMPK----- 198
Db 234 GGLFEVAGAMICKLWERTLGAIVROKNHMPTEAVKANKKICDEFENASKPOSIOESTG 233
OY 199 -----PAHTA----- 204
Db 294 SIEVLKIDSEGVASANTSRATSTSGFAGAIQKLPFSYATLELAIMYALGVGA 353
OY 205 -----AMDLSG-----POIESQCRV----- 221
Db 354 SIKDPKDLFTYEGSSDESCLPTEGVIIIGQSKMGGGLAEIPGLSINFAVLHGEQYLEL 413
OY 222 -----DIADA-----AYSIFOK----- 224
Db 414 YKPLPAGLKCACVAVADVLDKSGSVIIMDYVSYSEKELICHNOPSFLVSGSGGKR 473
OY 235 -----PKSFTGNFV-----IDENILKE 251
Db 474 TSDKVVAVAIIPRPDAVLTDTTSLNOALYRLSGDMNPLHIDPWFASLAGFDPKIL- 531
OY 252 EGIENE-----DVAIRK-----PGHLPDPDFLEDEPEAVSKYVE 286
Db 532 HGLCTGFESARVYLOQFADNDVSRFAIKARAKPVPYEGOTLOTETEMKGNRIHPTKVQ 591
OY 287 STGAVEPEFEKQLQIP-----KPRSGAVEEFT-----RIYVDSLDDVYKATQAI 332
Db 592 ETCGD-VISNAVYDLAPTSGETSAKTPSEGGKLOSTFVEEIRGLKD-IGPEVVKVNAV 649
OY 333 YLEFLS--GEDGTWFLDLKSKGNGVGEPSQADVWSMTTDDVVKMFSGKLAKTMAF 390
Db 650 FEWHITKGNIGAKWITIDIKSGSKYQGPAGKADTTIILDEDEMEVVLKLDPOKAF 709
OY 391 MSGKTKINGNMALAKLEKLM 411
Db 710 FSGRLKARGNIMLSOKLQML 730

RESULT 5
YALA_RHISN
ID YALA_RHISN STANDARD: PRT: 278 AA.
AC P55541:
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 01-NOV-1997 (rel. 35, Last annotation update)
DE PUTATIVE SHORT-CHAIN TYPE DEHYDROGENASE/REDUCTASE YALA (EC 1.-.-.-).
GN YALA.
OS Rhizobium sp. (strain NGR234).
OC Plasmid sym PNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-97305956; PubMed-9163424;
RA Fretberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RT Nature 387:394-401(1997).
RL Nature 387:394-401(1997).
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AE000082; AAB91754.1; -
CC HSP: P29132; IDP1.

```

```

DR InterPro: IP002398; -
DR InterPro: IP002347; -
DR Pfam: PF00106; adh_shortc.1.
DR Pfam: PF00678; adh_shortc2.1.
DR PRINTS: PR00080; SDRFAMILY.
DR PRINTS: PR00081; GDRHDH.
DR PROSITE: PS00061; ADH_SHORT.1.
KW Hypothetical protein; Oxidoreductase; Plasmid.
FT NP_BIND 10 34 NAD OR NADP (BY SIMILARITY).
FT ACT_SITE 159 159 BY SIMILARITY.
SQ SEQUENCE 278 AA; 28743 MW; 1D0105625DE9DF2B CRC64;

Query Match 10.8%; Score 232; DB 1; Length 278;
Best Local Similarity 29.5%; Pred. No. 16-09;
Matches 80; Conservative 34; Mismatches 107; Indels 50; Gaps 8;

OY 7 RLACTVFITGASRGIGKAIKAKNDGANIYA-----AKTQPPKLLGTITTA 58
Db 3 REEGKVAAYVYGAAGIGKACALAIAREGRRVYVADGSAIACTAO----- 49
OY 59 EETEAAGKALPCIVDVRDEQISAAVEAKIKKFGIDILVNNASISLT-----NTLDP 114
Db 50 --IAAEGNALAMADIADQAVALFETAEHFGVDLLVNNASMHILTPDRALDLD 107
OY 115 TKRLDLMNVNTRGYLASKACIPYLKSKVAHILNISPLNLPVPRQHCAYTIATG 174
Db 108 LAWMDQTMATNLGCTLCCROALPRMIARGGAIYVMSSCQGIS--GDPAQTSYAVSKAA 165
OY 175 MSNIVYGMAEF-KGEIYVNALMPKTAI-----HTAAMDLSGSGIESQGR 219
Db 166 MNNLSASLATQIGHAORICAAVAPGLMTERLAKDKCMQRHLRHQLPRVG-----H 220
OY 220 KVDIIADAAVSIFQKPKSFTGNFV-IDENIL 249
Db 221 PEDVVALVAFLLSDDSFTTGQVYCIDGGWL 251

RESULT 6
FOX2_NEUCR
ID FOX2_NEUCR STANDARD: PRT: 894 AA.
AC 001373:
DT 15-DEC-1998 (rel. 37, Created)
DT 15-DEC-1998 (rel. 37, Last sequence update)
DT 30-MAY-2000 (rel. 39, Last annotation update)
DE PEROXISOMAL HYDRAVASE-DEHYDROGENASE-EPIMERASE (HDE) (MULTIFUNCTIONAL
DE BETA-OXIDATION PROTEIN) (MFP) (INCLUDES: 2-ENOYL-COA HYDRATASE
DE (EC 4.2.1.-); D-3-HYDROXYACYL COA DEHYDROGENASE (EC 1.1.1.-)).
GN FOX-2.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 497-525 AND 604-619.
RX STRAIN-74-OR8-1A / DSM 1258;
RX MEDLINE-95231521; PubMed-7715608;
RA Fossa A., Beyer A., Pfister E., Wenzel B., Kunau W.-H.;
RT "Molecular cloning, sequencing and sequence analysis of the fox-2 gene
RT of Neurospora crassa encoding the multifunctional beta-oxidation
RT protein.";
RL Mol. Gen. Genet. 247:95-104(1995).
CC -1- FUNCTION: SECOND TRIFUNCTIONAL ENZYME ACTING ON THE BETA-OXIDATION
CC PATHWAY FOR FATTY ACIDS, POSSESSING HYDRATASE-DEHYDROGENASE-
CC EPIMERASE ACTIVITIES. CONVERTS TRANS-2-ENOYL-COA VIA D-3-
CC HYDROXYACYL-COA TO 3-KETOACYL-COA.
CC -1- PATHWAY: BETA-OXIDATION PATHWAY.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CATALASE-FREE MICROBODIES.
CC -1- DOMAIN: CONTAINS TWO SDR DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X80052; CA56355.1; -
 DR InterPro: IPR002198; -
 DR InterPro: IPR002347; -
 DR InterPro: IPR002539; -
 DR Pfam: PF01575; Maoc_denydratas; 1.
 DR Pfam: PF00106; adh_short; 2.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PRINTS: PR00081; GDRHD.
 DR PROSITE: PS00061; ADH_SHORT; 1.
 KW Fatty acid metabolism; Multifunctional enzyme; Oxidoreductase; NAD;
 KW Lyase; Isomerase; Repeat.
 FT DOMAIN 6 230 SHORT-CHAIN DEHYDROGENASE LIKE.
 FT NP_BIND 13 523 NAD (BY SIMILARITY).
 FT ACT_SITE 164 164 BY SIMILARITY.
 FT NP_BIND 318 342 NAD (BY SIMILARITY).
 FT ACT_SITE 458 458 BY SIMILARITY.
 SQ SEQUENCE 894 AA; 96326 MW; 29758220D397AA88 CRC64;

Query Match 10.8%; Score 231; DB 1; Length 894;
 Best Local Similarity 24.2%; Pred. No. 5,6e-09;
 Matches 114; Conservative 64; Mismatches 167; Indels 126; Gaps 20;

QY 7 RLAAGTAVITGASRQITGAIAIAKAKDGANIYI-----AAKTAQPHKILGTI 54
 DB 6 RRDGGVAVVTVGGGGLGAYACIFGSGRGASVYVNDLGASFKEGSGSTKAD----- 56
 QY 55 YTAEEIEAVGKALPCIVDVRDEQOISAIVEKAIKFKGIDILVNNNSA-----ISLTNTL 111
 DB 57 -VVVEIEIAAGKAAVANDSVENGDKI---ITAIKEGRIIDLINNGCILRDISFKNK 112
 QY 112 DTPTRRLDMNVNRTGYTLYSKACIPYLKSKVAHILNISPLNL--NPVFKQHCAYT 169
 DB 113 D---EDMDLIEFVHVKGSKYKTAARAMPYFRKQKFGRVINTASAGLFGN---FGQ-ANYS 165
 QY 170 IAKYMSMYVLGMAEE-FKGEIAVNAALPKTAIHTAADMCGPGIESCKRVVDIADRA 228
 DB 166 AAKLGAVGFTETLAKEGKYNIISNVIAPIAA--SRMTETVMPDULALMKEMVPLVA 223
 QY 229 YSIFQPKPSFGNFV-----IDEN-----IKKEGIEHNDVYA 261
 DB 224 VLVHKNNSTSEGSIEFVGGGHHVAKLRWERSGGLLKADSEYTPGAILIKMDVDYDPS--- 280
 QY 262 IKRPHLOPDEFLDEYPEAVSKKVESTGAVPEFKEE-----KLQLOPPRSGA 309
 DB 261 -NPQPTGPNDFLALLESIKLGPNDPGEKVDKGRVALVTGGAGIGAYLAFARAGA 339
 QY 310 VEETRIYKDSIS--DDVYKATQATITLFEISGEDGTWFLDKSKSGN--VGCEPSDQADV 367
 DB 340 SV-----VVNDLVNPDVVN-----EIKKMGKAVGAFKFSLEDGDA 375
 QY 368 VMSMTDFVKKFSGKLPYMAFMGSKLIRKGNMALIKLEKLMQNMARL 418
 DB 376 VKKAIDAF-----GRVDIVVNA--GILRDKAFINMDSL 409

RESULT 7
 ID UCPI_ECOLI STANDARD: PRT: 285 AA.
 AC P37440; P77442; P76863; P77140;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE OXIDOREDUCTASE UCPIA (EC 1.-.-.-).

GN UCPIA.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=97349980; PubMed=9205837;
 RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsunashi N.,
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
 RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubaram S.,
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
 RA Yamagata S., Horiuchi T.;
 RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
 RT K-12 genome corresponding to 50,0-68.8 min on the linkage map and
 RT analysis of its sequence features."
 RL DNA Res. 4:91-113(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=97385354; PubMed=9241368;
 RA Sirko A., Weglenska A., Hryniewicz M.M., Hulanicka D.M.;
 RT "Characterization of the Escherichia coli gene encoding a new member
 RT of the short-chain dehydrogenase/reductase (SDR) family."
 RL Acta Biochim. Pol. 44:153-157(1997).
 RN [4]
 RP SEQUENCE OF 202-285 FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=90264335; PubMed=218959;
 RA Hryniewicz M.M., Sirko A., Paluch A., Boeck A., Hulanicka D.M.;
 RT "Sulfate and thiosulfate transport in Escherichia coli K-12:
 RT identification of a gene encoding a novel protein involved in
 RT thiosulfate binding."
 RL J. Bacteriol. 172:3358-3366(1990).
 RN [5]
 RP IDENTIFICATION.
 RX MEDLINE=95075659; PubMed=7984428;
 RA Borodovsky M., Rudd K.E., Koonin E.V.;
 RT "Intrinsic and extrinsic approaches for detecting genes in a
 RT bacterial genome."
 RL Nucleic Acids Res. 22:4756-4767(1994).
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC (SDR) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE000330; AAC75479.1; -
 DR EMBL: D90872; BAA16309.1; -
 DR EMBL: D90871; BAA16300.1; -
 DR EMBL: X99908; CAN68181.1; ALT_INIT.
 DR EMBL: M32101; -, NOT_ANNOTATED_CDS.
 DR HSSP: P14061; IFDV.
 DR Ecogene: EG12133; ucpiA.
 DR InterPro: IPR002198; -
 DR InterPro: IPR002347; -


```
RN [4]
RP SEQUENCE OF 230-246 FROM N.A.
RC STRAIN-168;
RA Oguro A., Takeshita H., Takamatsu H., Nakamura K., Yamane K.;
RA Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: (3R)-3-HYDROXYACYL-[ACYL-CARRIER PROTEIN] +
CC NADP(+) -> 3-OXOACYL-[ACYL-CARRIER PROTEIN] + NADPH.
CC -1- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
CC PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U59433; AAC44307.1; -
CC EMBL: Z59112; CAB13464.1; -
CC EMBL: Y13937; CAA74250.1; -
CC EMBL: D64116; BAA10974.1; -
CC HSSP: Q12634; 1YBV.
CC Subtilist: BG11535; fabG.
CC InterPro: IPR002198; -.
CC InterPro: IPR002347; -.
CC Pfam: PF00106; adh_short.1.
CC Pfam: PF00678; adh_short.C2; 1.
CC PRINTS: PR00080; SDRFAMILY.
CC PRINTS: PR00081; GDRHD.
CC PROSITE: PS00061; ADH_SHORT; FALSE_NEG.
CC Fatty acid biosynthesis; Oxidoreductase; NADP.
CC NP_BIND 9 NADP (BY SIMILARITY).
CC ACT_SITE 154 154 BY SIMILARITY.
CC CONFLICT 23 23 D -> A (IN REF. 1).
CC SEQUENCE 246 AA; 26282 MW; C6A391167D3237DC CRC64;

Query Match 10.2%; Score 218.5; DB 1; Length 246;
Best Local Similarity 31.0%; Pred. No. 7.8e-09;
Matches 63; Conservative 36; Mismatches 63; Indels 41; Gaps 8;

QY 12 TVFTTASRGICGKALIAKAKDGANIVT-----AAKTAQHPKLLCTITYTAAEIEAVG 66
D 6 TAIYTGASRGISIALDLAKSGANVYVNSGNEAKANE-----VVDKIKSMGR 54
QY 67 KALPCIVDRDEQOISAAREKAKKFGIDILVNNASAIISLTNTLDPTPKRL-----DLM 121
D 55 KALVAVADVSNPEVDONMKETILSVSTIDILVNNAGT-----TRDNLIMRKEDMDV 109
QY 122 MANNTRGTYLASKACIPYLKSKVAHILNISPLNL--NPWFKQHCAYTIAKGSMSV 179
D 110 ININLKGVFNCRTKAVTRQMMKQKRSRIINSSIVGSGNP-----GGANYVAANAG-----Y 161
QY 180 LGMAEEFKGEIA-----VNAAMP 197
D 162 IGLTKSSAKELASRNITVNAIAP 184

RESULT 12
Y325_THEMEA
ID Y325_THEMEA STANDARD; PRT; 251 AA.
AC Q9WYG0;
DT 30-MAY-2000 (rel. 39, Created)
DT 30-MAY-2000 (rel. 39, Last sequence update)
DT 30-MAY-2000 (rel. 39, Last annotation update)
DE HYPOTHETICAL OXIDOREDUCTASE TM0325 (EC 1.-.-.-).
GN TM0325.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MSB8 / DSM 3109;
RA MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Uitterback T.R., Malek J.A., Lihner K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RL genome sequence of Thermotoga maritima."
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR TIGR: TM0325; AAD35412.1; -
CC InterPro: IPR002198; -.
CC InterPro: IPR002347; -.
CC InterPro: IPR002424; -.
CC Pfam: PF00106; adh_short.1.
CC Pfam: PF00678; adh_short.C2; 1.
CC PRINTS: PR00080; SDRFAMILY.
CC PRINTS: PR00081; GDRHD.
CC PRINTS: PR01167; INSADHFAMILY.
CC PROSITE: PS00061; ADH_SHORT; 1.
CC Hypothetical protein; Oxidoreductase.
CC NP_BIND 10 NADP (BY SIMILARITY).
CC ACT_SITE 152 152 BY SIMILARITY.
CC SEQUENCE 251 AA; 26635 MW; 61C736A0F13564A5 CRC64;

Query Match 10.2%; Score 218.5; DB 1; Length 251;
Best Local Similarity 30.7%; Pred. No. 8e-09;
Matches 66; Conservative 38; Mismatches 96; Indels 15; Gaps 6;

QY 10 GCTVFTTASRGICGKALIAKAKDGANIVTAAKTAQHPKLLCTITYTAAEIEAVGKAL 69
D 5 GKVVLTITGASGSGKKAAMFAERGAKVAINDISSEKKE-----TYELIKSMGEAA 57
QY 70 PCIVDV-RDEQOISAAREKAKKFGIDILVNNASAIISLTNTLDPTPKRLDLMANNVTRG 128
D 58 FIFGDVAKDAEQI---VKRTVEFGRDLVNNAGIVPYGNIETSEEDFDKTMANNVNG 114
QY 129 TYLASKACIPYLKSKVAHILNISPLNLNPWFKQHCAYTIAKGSMSVYVGMMAEEFKG 188
D 115 PFLSKYAVEQMKKGGGVIYVNSSEAGL--IGIPRCVYSKAAALGLTSLSLAVDYD 172
QY 189 -ETAVNALMPKTAIHTPAMDML-GGPGIESQCRKV 221
D 173 YGIRVNAVCPGTTQSEGLMARVAKASPNPELLKKM 207

RESULT 13
DHG_BACSU
ID DHG_BACSU STANDARD; PRT; 261 AA.
AC P12310; P94430;
DT 01-OCT-1989 (rel. 12, Created)
DT 30-MAY-2000 (rel. 39, Last sequence update)
DT 30-MAY-2000 (rel. 39, Last annotation update)
DE GLUCOSE 1-DEHYDROGENASE (EC 1.1.1.47).
GN GDH.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
```


AC 002207;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE PROXISOMAL HYDRATASE-DEHYDROGENASE-EPIMERASE (HDE) (MULTIFUNCTIONAL
 BETA-OXIDATION PROTEIN) (MFP) [INCLUDES: 2-ENOYL-COA HYDRATASE
 (EC 4.2.1.-); D-3-HYDROXYACYL COA DEHYDROGENASE (EC 1.1.1.-)].
 GN FOX2 OR YKR009C OR YK108.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=93070612; PubMed=1441752;
 RA Diesterhoef A., Philippaen P.;
 RT "DNA sequencing and analysis of a 24.7 kb segment encompassing
 RT centromere CEN1 of Saccharomyces cerevisiae reveals nine previously
 RT unknown open reading frames.";
 RL Yeast 8:749-759(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92202210; PubMed=1551874;
 RA Hiltunen J.K., Wenzel B., Beyer A., Erdmann R., Fossa A., Kunau W.H.;
 RT "Peroxisomal multifunctional beta-oxidation protein of Saccharomyces
 RT cerevisiae. Molecular analysis of the fox2 gene and gene product.";
 RL J. Biol. Chem. 267:6646-6653(1992).
 CC -1- FUNCTION. SECOND TRIFUNCTIONAL ENZYME ACTING ON THE BETA-OXIDATION
 CC PATHWAY FOR FATTY ACIDS, POSSESSING HYDRATASE-DEHYDROGENASE-
 CC EPIMERASE ACTIVITIES. CONVERTS TRANS-2-ENOYL-COA VIA D-3-
 CC HYDROXYACYL-COA TO 3-KETOACYL-COA.
 CC -1- PATHWAY: BETA-OXIDATION PATHWAY.
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: PEROXISOMAL.
 CC -1- DOMAIN: CONTAINS TWO SDR DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC (SDR) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M86456; AAA34779.1; -;
 DR EMBL: X65124; CAA46243.1; -;
 DR EMBL: Z28234; CAA82079.1; -;
 DR PIR: S25322; S25322.
 DR HSSP: P25529; 1AHH.
 DR SGD: S0001717; FOX2.
 DR InterPro: IPR002198; -;
 DR InterPro: IPR002539; -;
 DR Pfam: PF01575; Maoc_dehydratase; 1.
 DR Pfam: PF00106; adh_short; 2.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PROSITE: PS00061; ADH_SHORT; 2.
 DR PROSITE: PS00342; MICROBODIES_CTER; 1.
 KW Fatty acid metabolism; Multifunctional enzyme; Oxidoreductase; NAD;
 KW Lyase; Isomerase; Peroxisome; Repeat.
 FT DOMAIN 6 230 SHORT-CHAIN DEHYDROGENASE LIKE.
 FT SITE 319 535 SHORT-CHAIN DEHYDROGENASE LIKE.
 FT SITE 898 900 MICROBODY TARGETING SIGNAL (POTENTIAL).
 FT NP_BIND 13 37 NAD (BY SIMILARITY).
 FT ACT_SITE 165 165 NAD (BY SIMILARITY).
 FT NP_BIND 326 350 NAD (BY SIMILARITY).
 FT ACT_SITE 469 469 BY SIMILARITY.
 FT ACT_SITE 900 AA; 98703 MM; 66FDD049C673788 CRC64;
 SO SEQUENCE

Query Match 9.98; Score 211; DB 1; Length 900;

Best Local Similarity 26.08; Pred. No. 1,5e-07;
 Matches 109; Conservative 50; Mismatches 147; Indels 114; Gaps 21;
 QY 13 VFTIGASRGICATATLAAKDGANIVIAKTAQHPPLTGTI-----YTAEEIE 62
 DB 12 VVTGAGGGLKVALALVASKAVVY-----NDLGTIGSGGSKNAADLYDEIK 63
 QY 63 AVGKALPCIVDVDEQOISAVERKAIRKFGIDILVNNASAIJLNTLDTPTKRLDLM 122
 DB 64 KAGGIAVANVDSVNGE--KIETAIKEGVRVLLNNGILRDVSFAKWTREFAVY 121
 QY 123 NVNTRGTYLASKACIPYLKSKVAHILNISPLNL--NPWFKORCAVTTAKGMSHYVL 180
 DB 122 DVHLTGKYLKSRAMPYMRSGKFGRIINTASPAGLFEN---FGQ-ANYSAKMGV---V 173
 QY 181 GMAEEF-----KGEIYVNALMPKTAIHTAAMDMLGGPISQCKKVDIIDAAY----- 229
 DB 174 GLAETTLAKKGAQKINYNVSLAP--LANSKRTENVLPHELKQLGPEKIVPLVLYLTHEST 231
 QY 230 ---SIFQPKSFTGN-----FVIDENILKEGIEIEN---FDVYAIKPGHPLOPD 271
 DB 232 KVSNSIFELAGFEGQLRMERSGQIFNPDKTYTPPAILNKKWEITDYDRKPFNKTOHP 291
 QY 272 FFLDEYEPVASKVESGAVPEEKEKLOPKRSGAVETFRIVVDSLSDDVYKATQA 331
 DB 292 YQLSDYNDLIT-----RAKRL---PPNEGQSVK---IK-SLCNKVYVY--- 328
 QY 332 IYLFELSGEDG-----TWFL-----DLKSKG-----NVGGEPS---DQADV 368
 DB 329 -----GAGGGLCKSHALWPARYGARVYVNDIDPSSVYEIKKLKGEGALIPDSHDV 381

Search completed: June 11, 2001, 16:25:47
 Job time: 137 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2001, 16:20:50 ; Search time 22.06 Seconds
(without alignments)
1063.147 Million cell updates/sec

Title: US-09-464-039-7

Perfect score: 2139
Sequence: 1 MLPNTGRLAGCTVFITGASR.....GNMALIKLEKLNQNMARL 418

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq_0401.*
1: /SID56/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SID56/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SID56/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SID56/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SID56/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SID56/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SID56/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SID56/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SID56/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SID56/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SID56/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SID56/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SID56/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SID56/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SID56/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SID56/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SID56/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SID56/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SID56/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SID56/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SID56/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SID56/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	244	11.4	258	21	Bacillus D-arabini
2	232.5	10.9	735	18	Human host cell pr
3	224	10.5	240	16	Clavulanic acid de
4	218.5	10.2	247	16	S. clavuligerus OR
5	217	10.1	248	21	Amino acid sequenc
6	217	10.1	261	21	B. subtilis glucos
7	217	10.1	261	21	Amino acid sequenc
8	206.5	9.7	900	20	A multifunctional
9	206.5	9.7	315	17	Rape leaf beta-ket
10	203.5	9.5	313	21	Rape seed beta-ket
11	203.5	9.5	313	21	Human ORFX ORF2404

12	202.5	9.5	313	20	Y27004	Human short-chain
13	199.5	9.3	261	13	R27757	Glucose dehydrogen
14	198.5	9.3	261	11	R03846	Modified glucose d
15	198	9.3	251	21	Y55986	S. clavuligerus cla
16	197.5	9.2	903	20	Y28713	Detargeted yeast m
17	197.5	9.2	906	20	Y28702	Yeast multifunctio
18	197.5	9.2	906	20	Y28712	Mutant yeast multi
19	197	9.2	255	13	R22983	7-alpha-hydroxy st
20	196.5	9.2	261	11	R04044	Glucose dehydrogen
21	196.5	9.2	261	13	R24018	Thermostable gluco
22	193.5	9.0	285	21	G40333	Arabidopsis thalia
23	193.5	9.0	307	21	G40332	Arabidopsis thalia
24	191.5	9.0	246	21	B15707	Staphylococcus aur
25	191.5	9.0	262	20	Y08339	A. parasiticus ver
26	191.5	9.0	298	21	G50312	Arabidopsis thalia
27	191.5	9.0	315	21	G50311	Arabidopsis thalia
28	190.5	8.9	298	21	G24787	Arabidopsis thalia
29	190.5	8.9	316	21	G24786	Arabidopsis thalia
30	190	8.9	285	21	G12862	Arabidopsis thalia
31	190	8.9	307	21	G12861	Arabidopsis thalia
32	189.5	8.9	186	19	W38474	S. pneumoniae 3-ox
33	189.5	8.9	243	19	W80670	S. pneumoniae fat
34	189.5	8.9	243	21	B15706	Streptococcus pneu
35	189	8.8	272	21	B10740	B. megaterium gluc
36	189	8.8	340	21	B10741	H. ghilianii/B. me
37	188.5	8.8	292	21	Y98000	Human SCAD family
38	186.5	8.7	260	17	R94617	3-hydroxybutyric a
39	186	8.7	261	9	P80590	Sequence of glucos
40	186	8.7	261	13	R27756	NAD affinity glucos
41	185.5	8.7	277	21	B29473	Burkholderia sp. C
42	185	8.6	261	9	P80063	Glucose dehydrogen
43	183	8.5	276	16	R66734	Aromatic dihydrodi
44	181.5	8.5	285	21	G12202	Arabidopsis thalia
45	181.5	8.5	285	21	G47575	Arabidopsis thalia

ALIGNMENTS

RESULT 1	
Y56815	Y56815 standard; Protein: 258 AA.
XX	
AC	Y56815;
XX	
DT	31-MAR-2000 (first entry)
XX	
DE	Bacillus D-arabinitol dehydrogenase.
XX	
KW	D-arabinitol dehydrogenase; clinical diagnosis; mycosis.
XX	
OS	Bacillus sp.
XX	
PN	JPL1332569-A.
XX	
PD	07-DEC-1999.
XX	
PF	26-MAY-1998; 98JP-0143637.
XX	
PR	26-MAY-1998; 98JP-0143637.
XX	
PA	(IKED-) IKEDA SHOKKEN KK.
XX	
PA	(NIPK) NIPPON KAYAKU KK.
XX	
DR	WPI: 2000-091353/08.
XX	
DR	N-PSDB: Z46762, Z46763.
XX	
PT	Arabidinitol dehydrogenase gene encoding D-arabinitol dehydrogenase -
XX	
PS	useful as a clinical diagnosing agent for mycosis
XX	
XX	Claim 1; Page 10-11; 14pp; Japanese.
CC	The invention relates to gene encoding D-arabinitol dehydrogenase,


```

DE Clavulanic acid dehydrogenase sequence.
XX
XX Clavulanic acid; antibiotic; Augmentin.
XX
XX Streptomyces clavuligerus ATCC 27064.
XX
XX WO9503416-A.
XX
XX 02-FEB-1995.
XX
XX 15-JUL-1994; 94WO-EP02346.
XX
XX 24-JUL-1993; 93GB-0015393.
XX
XX (SMIK ) SMITHKLINE BEECHAM PLC.
XX
XX Arnell J, Elson SW, Nicholson NH, Woronicki SR;
XX
XX WPI: 1995-075242/10.
XX
XX N-PSDB; Q67531.
XX
XX New clavulanic acid dehydrogenase from Streptomyces, and related
XX DNA and vectors - used to produce beta-lactamase inhibiting
XX clavulanic acid from new 3-oxo:ethylidene analogues
XX
XX Claim 2; Fig 1; 36pp; English.
XX
XX A new enzyme is disclosed which has clavulanic acid dehydrogenase
XX activity, the enzyme having an apparent mol. wt. of 28 kD (by SDS
XX PAGE) and incorporating the N-terminal sequence PSALGKVALRIGASSGIE.
XX The enzyme is derived from the mycelium of a Streptomyces species,
XX especially S. clavuligerus (e.g. ATCC 27064), S. jumigensis (e.g. ATCC
XX 29864) or S. katushahannus (e.g. 7272). The present sequence
XX represents the enzyme from S. clavuligerus ATCC 27064.
XX The enzyme is used to catalyse the biosynthesis of clavulanic acid
XX from a precursor clavulanic acid aldehyde. The obtained clavulanic acid
XX is in turn a key ingredient in the antibiotic Augmentin.
XX
XX Sequence 240 AA:
SQ
Query Match 10.5%; Score 224; DB 16; Length 240;
Best Local Similarity 29.5%; Pred. No. 1.3e-13;
Matches 71; Conservative 40; Mismatches 96; Indels 34; Gaps 8;
OY 8 LAGCVFTTGSRGKAIKAKDGANIVIAKTAQPHPKLGITTAAEITAVGSK 67
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5 IqkvvailltgassgIgeataraIaegaavaIaarrve-----kIralgdelIaaagk 57
OY 68 ALPCIVDVRDEQOISAAVEKAIRKFGGIDILVNNASAI SLTNTLDTPTKRIDLMMNVNTR 127
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 58 vhlvdvdrdqgvdavaavsteaIgdilIvnaagImlIgpvedadtwtmIdtlnl 117
OY 128 GTYLAKKACIPYLKSK--VAHILNISPPILNIPWFKQHCAYTIAKGYMSYVLGMAEE 185
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 118 glmymtiraalphlIrlskgtlvqmsIagrIv-----fnaavYgatkIgvnaIafsetIqge 172
OY 186 F-KGEIANNALMPKTA-----IHTAADMVLGSGPIE---SOCRAK---DIADAAYS 230
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 173 vtergvrvvIepgttdelIghlthtakem-----YegrlsgrIkIqadqlaeavrya 227
OY 231 I 231
| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 228 v 228

```

```

XX
XX S. clavuligerus ORF9 product.
DE
XX Clavulinic acid; clavulinate; antibiotic; beta-lactamase-inhibitor.
XX
XX Streptomyces clavuligerus.
XX
XX CA2108113-A.
XX
XX 09-APR-1995.
XX
XX 08-OCT-1993; 93CA-2108113.
XX
XX 08-OCT-1993; 93CA-2108113.
XX
XX (UYAL-) UNIV ALBERTA.
XX
XX AIdoo KA, Jensen SE, Paradkar AS;
XX
XX WPI: 1995-207301/28.
XX
XX N-PSDB; Q91580.
XX
XX Clavulanic acid biosynthesis enzymes and corresp. DNA - useful for
XX biosynthesis of the antibiotic in Streptomyces hosts which do not
XX naturally produce clavulanate
XX
XX Claim 31; Fig.18; 41pp; English.
XX
XX A 15 kb fragment S. clavuligerus NRRL 5741 genomic DNA (Q91580),
XX encoding downstream from pcbc, included 10 ORFs encoding the
XX enzymes required for clavulanate biosynthesis. ORF8 (oriented
XX in the opposite direction to pcbc) encoded a protein (R77866)
XX that showed weak similarity to ribitol 5 P04 dehydrogenase-type
XX enzymes.
XX
XX Sequence 247 AA:
SQ
Query Match 10.5%; Score 224; DB 16; Length 247;
Best Local Similarity 29.5%; Pred. No. 1.4e-13;
Matches 71; Conservative 40; Mismatches 96; Indels 34; Gaps 8;
OY 8 LAGCVFTTGSRGKAIKAKDGANIVIAKTAQPHPKLGITTAAEITAVGSK 67
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5 IqkvvailltgassgIgeataraIaegaavaIaarrve-----kIralgdelIaaagk 57
OY 68 ALPCIVDVRDEQOISAAVEKAIRKFGGIDILVNNASAI SLTNTLDTPTKRIDLMMNVNTR 127
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 58 vhlvdvdrdqgvdavaavsteaIgdilIvnaagImlIgpvedadtwtmIdtlnl 117
OY 128 GTYLAKKACIPYLKSK--VAHILNISPPILNIPWFKQHCAYTIAKGYMSYVLGMAEE 185
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 118 glmymtiraalphlIrlskgtlvqmsIagrIv-----fnaavYgatkIgvnaIafsetIqge 172
OY 186 F-KGEIANNALMPKTA-----IHTAADMVLGSGPIE---SOCRAK---DIADAAYS 230
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 173 vtergvrvvIepgttdelIghlthtakem-----YegrlsgrIkIqadqlaeavrya 227
OY 231 I 231
| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 228 v 228

```

```

RESULT 4
ID R77866 standard; Protein; 247 AA.
XX
AC R77866;
XX
DT 13-NOV-1995 (first entry)

```

```

RESULT 5
ID Y54422 standard; Protein; 248 AA.
XX
AC Y54422;
XX
DT 06-APR-2000 (first entry)
XX
XX Amino acid sequence of a beta-ketoacyl-ACP reductase protein.
XX

```

KV	Beta-ketoacyl-ACP reductase; Type II fatty acid synthetase;
KW	stereoselectivity; 4-chloroacetoacetyl acid ester;
KM	(S)-4-halo-3-hydroxybutyric acid ester; 4-halo-acetoacetic acid ester;
KK	beta-ketoacyl-acyl carrier protein reductase; acetoacetyl-CoA reductase;
KW	poly(beta-hydroxy fatty acid biosynthesis); optically active;
KM	4-naol-3-hydroxybutyric acid ester.
XX	
OS	Bacillus subtilis.
PN	EP955375-A2.
PD	10-NOV-1999.
XX	
PE	10-MAY-1999; 99EP-0109403.
XX	
PR	08-MAY-1998; 98JP-0126507.
PR	21-OCT-1998; 98JP-0300178.
PR	05-APR-1999; 99JP-0098205.
XX	
PA	(DAIL) DAICEL CHEM IND LTD.
PI	Yamamoto H;
XX	
DR	WPI; 2000-118183/11.
DR	N-PDB; 245749.
XX	
PT	Producing optically pure (S)-4-halo-3-hydroxybutyric acid ester -
XX	
PS	Claim 6; Page 19-20; 34pp; English.
CC	The present sequence represents a beta-ketoacyl-ACP reductase protein
CC	of Bacillus subtilis. The beta-ketoacyl-ACP reductase enzyme constitutes
CC	a type II fatty acid synthetase. The enzyme has an extremely high
CC	reducing activity and stereoselectivity towards 4-chloroacetoacetyl
CC	acid ester. The specification describes a method for producing a
CC	(S)-4-halo-3-hydroxybutyric acid ester. The method comprises
CC	asymmetrically reducing 4-halo-acetoacetic acid ester or its
CC	derivative with beta-ketoacyl-acyl carrier protein reductase
CC	constituting Type II fatty acid synthase, or acetoacetyl-CoA
CC	reductase constituting the poly(beta-hydroxy fatty acid biosynthesis
CC	system. The novel method is used to produce optically active
CC	4-naol-3-hydroxybutyric acid ester, with a high purity.
SO	Sequence 248 AA:
Query Match	10.2%; Score 218.5; DB 21; Length 248;
Best Local Similarity	31.0%; Pred.No.4.8e-13;
Matches 63; Conservative 36; Mismatches 63; Indels 41; Gaps 8;	
OY	12 TVEFTGASGTCIKATLAKAKGANIVT-----AAKTQPHFKLLGTITTAEELEAVAGG 66 : ::: : : : :
Db	8 talvagsargtigrsalakalsaganvvnyvsneake-----vvdeiksmgr 56 : : : : : : : : : : : : :
OY	67 KALPCIVDPDRDOQSANAEKAIRKFEGGDILVNNASALSINTLTDTPTKR L-----DLM 121 : : : : : : : : : : : : :
Db	57 kalavkadvsnpedqymmlkelvsvstldilvmnag1-----tcdnlmrkedegwdv 111 : : : : : : : : : : : : :
OY	122 MNVNRGTVFLAKACIPYLKSKVAHILNISPLNI--NPVMFKQHCAVTIAKYGSMNV 179 : : : : : : : : : : : : : : :
Db	112 iinlnkgyfnctkavlrgmmkgsgrrllnvsslvgsnpr---gqanyvaakag----v 163 : : : : : : : : : : : :
OY	180 LGMAEFPKEGA-----VNALMP 197 : : : : : : : : : : : :
Db	164 igltksakelasmrltvnaiap 186 : : : : : : : : : : : :
RESULT	6
ID	Y96271
XX	Y96271 standard; Protein; 261 AA.
AC	Y96271.
XX	

```

DT      31-AUG-2000   (first entry)
XX
DE      B. subtilis' glucose dehydrogenase.
XX
KW      Carbonyl reductase; (S)-4-halo-3-hydroxybutyrate ester; drug synthesis;
KM      reduced nicotinamide adenine dinucleotide; NADH; alcohol production;
XX      4-haloacetoacetate ester; glucose dehydrogenase; enzyme.
OS      Bacillus subtilis.
XX      EP1013758-A2.
PN      28-JUN-2000.
PD      21-DEC-1999;    99EP-0125572.
PF      21-DEC-1999;    98JP-0363130.
XX      17-JUN-1999;    99JP-0171160.
PR      (DAIL ) DAICEL CHEM IND LTD.
PA
XX      Kimoto N, Yamamoto H, Mitsuhashi K;
PI      WPI; 2000-414601/36.
DR      N-PSDB; A38806.
XX
PT      New carbonyl reductase reduces 4-haloacetoacetate ester to produce
PT      (S)-4-halo-3-hydroxybutyrate ester, using reduced beta-nicotinamide
XX      adenine dinucleotide as an electron donor -
XX
PS      Example 20; page 28-29; 37pp; English.
XX
CC      Carbonyl reductase reduces 4-haloacetoacetate ester to produce
CC      (S)-4-halo-3-hydroxybutyrate ester, using reduced beta-nicotinamide
CC      adenine dinucleotide (NADH) as an electron donor. This enzyme is useful
CC      because it has a high reductase activity for 4-chloroacetoacetate ester
CC      but does not substantially dehydrogenate any optical isomers of
CC      4-halo-3-hydroxybutyrate ester. In comparison with other enzymes with
CC      similar activities, carbonyl reductase shows a higher enzymatic activity
CC      when used with reduced NADH than with beta-nicotinamide adenine
CC      dinucleotide phosphate (NADPH), which is expensive and chemically
CC      unstable. The reductase has excellent stereoselectivity producing
CC      optically active (S)-4-halo-3-hydroxybutyrate ester with high optical
CC      purity and in high yield. (S)-4-halo-3-hydroxybutyrate ester is a useful
CC      drug intermediate and is also of importance in alcohol production. The
CC      present sequence is the glucose dehydrogenase protein of Bacillus
CC      subtilis. Glucose dehydrogenase is important in the regeneration of
CC      NADPH. The glucose dehydrogenase gene (A38806) was inserted into a
CC      plasmid, already carrying the carbonyl reductase gene of Kluyveromyces
CC      fragilis resulting in coexpression of the two genes from the same
CC      plasmid.
XX
SQ      Sequence      261 AA;

```

DB 170 etlaIeyapkg-irvniIgp-gaint 193

RESULT 7

ID Y54424 standard; Protein: 261 AA.

AC Y54424;

DT 06-APR-2000 (first entry)

DE Amino acid sequence of the glucose dehydrogenase enzyme.

KM Glucose dehydrogenase gene: Type II fatty acid synthetase;

KW stereoselectivity; 4-chloroacetoacetic acid ester;

KM (S)-4-halo-3-hydroxybutyric acid ester; 4-halo-acetoacetic acid ester;

KM beta-ketoacyl-acyl carrier protein reductase; acetoacetyl-CoA reductase;

KM polybeta-hydroxy fatty acid biosynthesis; optically active;

OS Bacillus subtilis.

PN EP955375-A2.

PD 10-NOV-1999.

PF 10-MAY-1999; 99EP-0109403.

PR 08-MAY-1998; 98JP-0126507.

PR 21-OCT-1998; 98JP-0300178.

PR 05-APR-1999; 99JP-0098205.

PA (DAIL) DAICEL CHEM.IND LTD.

PI Yamamoto H;

DR WPI: 2000-118183/11.

DR N-PSDB: 245757.

PT Producing optically pure (S)-4-halo-3-hydroxybutyric acid ester -

PS Example 18; Page 27-28; 34pp; English.

XX The present sequence represents the glucose dehydrogenase protein of

CC Bacillus subtilis. The glucose dehydrogenase gene was cloned, in

CC the course of the invention, to regenerate the reduced nicotinamide

CC adenine dinucleotide phosphate. The specification describes a

CC method for producing a (S)-4-halo-3-hydroxybutyric acid ester.

CC The method comprises asymmetrically reducing 4-halo-acetoacetic

CC reductase constituting Type II fatty acid synthase, or acetoacetyl-CoA

CC reductase constituting the polybeta-hydroxy fatty acid biosynthesis

CC system. The novel method is used to produce optically active

CC 4-halo-3-hydroxybutyric acid ester, with a high purity.

XX Sequence 261 AA:

Query Match 10.1%; Score 217; DB 21; Length 261;

Best Local Similarity 31.6%; Pred. No. 7.2e-13;

Matches 65; Conservative 42; Mismatches 83; Indels 16; Gaps 7;

OY 1 MPEPTGRLAGCTVFITGASRGIGKALIAKADGANIVIAAKTAQPHKLLGTTTAAEE 60

DB 1 mypd--lkgyvaltgaaaglgkamalrfgkqkvvnylnyknqdpn-----evkee 51

OY 61 IAVVGKALPCIVDVEQOISAWEKAIAKFGGIDILVNNASISLTNTLDPTFRRLD 120

DB 52 vfkaggeavvvgdvtkedvknivqtaikefgcltdiminnaglenpyvshempkdkwk 111

OY 121 MNMVTGRTYLASKACIPYLKRSKV-AHLINISPLNLNPFVFKOHCAVTIAKYGSMYV 179

DB 112 vgltnltgtfsgrealkyfvendikgnvinnssvhevipwplfvh--yaaakgglkint 169

OY 180 LGMAEEF--KGEIYANALMPKTAIHT 203

DB 170 etlaIeyapkg-irvniIgp-gaint 193

RESULT 8

ID Y30338 standard; Protein: 900 AA.

AC Y30338;

DT 15-NOV-1999 (first entry)

DE A multifunctional enzyme designated Fox2.

KM Multifunctional enzyme: Fox2; plant metabolism; transgenic plant;

KM fatty acid oxidation enzyme: polyhydroxyalkanoate; oil composition;

OS Saccharomyces sp.

PN WO9945122-A1.

PD 10-SEP-1999.

PF 05-MAR-1999; 99WO-US04999.

PR 06-MAR-1998; 98US-0077107.

PA (META-) METABOLIX INC.

PI Boynton L, Huisman GW, Moloney M, Patterson N, Peoples OP;

DR WPI: 1999-540850/45.

DR N-PSDB: 210369.

PT Modifying fatty acid metabolism in plants, useful for increasing

PS biomass and producing specific polymers in seeds

XX Disclosure; Page 52-56; 79pp; English.

CC The present sequence represents a multifunctional enzyme designated Fox2.

CC Such multifunctional enzymes may be used in the method of the invention.

CC The specification describes a method for manipulating the metabolism of

CC a plant, and comprises expressing a heterologous gene encoding fatty acid

CC oxidation enzymes in the cytosol or plastids other than the peroxisomes,

CC glyoxisomes or mitochondria of the plant. The method may be used to

CC enhance the biological production of polyhydroxyalkanoates or novel oil

CC compositions in a transgenic plant. Plants which may be used to produce

CC these compounds in this way include Brassicas, maize, soybean,

CC cottonseed, sunflower, palm, coconut, safflower, peanut, mustard, flax,

CC tobacco and alfalfa. The method may also be used to prevent or suppress

CC seed production and therefore increase the production of biomass (leaves,

CC stems, stalks) by plants.

XX Sequence 900 AA:

Query Match 9.9%; Score 211; DB 20; Length 900;

Best Local Similarity 26.0%; Pred. No. 1.8e-11;

Matches 109; Conservative 50; Mismatches 147; Indels 114; Gaps 21;

OY 13 VFTGASRGIGKALIAKADGANIVIAAKTAQPHKLLGTTTAAEEIE 62

DB 12 vltgaggglgkvayalaysrsgakvv-----ndlgltlgsggshnskaadlvvdclk 63

OY 63 AVGKRALPCIVDVEQOISAWEKAIAKFGGIDILVNNASISLTNTLDPTFRRLDLM 122

DB 64 kagtlavanydsvnenge--kiletaikefgvrdvlinnagllrdvsfakmterefasv 121

OY 123 NNTNRTIYLASKACIPYLKRSKVVAHLINISPLNLN--NFWWFKOHCAVTIAKYGSMYV 180

```

Db      122 dvhltygylklsraawpymrsgkfgrlntaspaglfgn---fqq-aynsaakmg1-----v 173
OY      181 GMAEEF-----KGEIANNALMPKTAIHTAAMDMLGSGPIESQCRKVDIADAAV----- 229
Db      174 glaeetlkegagkylnvnslap--larsrmtenvlpplhllkqlspektvplvlylthest 231
OY      230 ----SIFQPKPSFTGN-----FVIDENILKEGCIEN----FDVYAIKPGHPLOPD 271
Db      232 kvansifeelaagffgqlrwerssglfnppdktytpaellnkwkeltidrydkfnkctqhp 291
OY      272 FFLDEYEAWSKVESGAVPEFEKEKLOLOPKRSGAVEETFRIVNDSLDVYKATQA 331
Db      292 yqlsdynlilt-----kakkl--pneqsgsvk-----lk-slcnkvvvvt-- 328
OY      332 IYLFELSGEDSG-----TWFL-----DLKSKG-----NVGGEPS-----DOADV 368
Db      329 -----gagggdglgshaiwfarvagkvvvndlkdpfsvveinkllygegtalpdshdv 381

```

RESULT 9

R89323
ID R89323 standard; Protein; 315 AA.

AC R89323;

DT 13-APR-1996 (first entry)

DE Rape leaf beta-ketoacyl-ACP-ketoreductase.

KW Rape; leaf; beta-ketoacyl-ACP-ketoreductase; beta-ketoreductase;
KM plasmid pJR16.2; CDNA library; Escherichia coli; vector;
KW plasmid; stroma; transit peptide; cassette; antisense; oilseed;
KM transgenic plant; crop improvement; lipid; metabolic engineering;
KM polymer; rapeseed oil.

OS Brassica napus.

FH Key Location/Qualifiers

FT Peptide 1..55 /note= "transit peptide"

PN W09602652-A2.

PD 01-FEB-1996.

PF 17-JUL-1995; 95WO-G801678.

PR 20-JUL-1994; 94GB-0014622.

PA (ZENE) ZENECA LTD.

PI Chase D, Elborough K, Fentem PA, Slabas AR, White A;

DR WPI: 1996-105914/11.

DR N-PSDB; Q99305.

PT New isolated rape beta-ketoreductase DNA - used to develop plants

PS Claim 2; Page 16; 29pp; English.

CC The sequence corresponds to a rape leaf beta-ketoreductase encoded
CC by a CDNA insert in plasmid pJR16.2 in Escherichia coli XL1-Blue.
CC A plasmid stroma targeting transit peptide is present. DNA
CC encoding the protein may be inserted in a vector or expression
CC cassette in sense or antisense orientation for expression in an
CC oilseed plant, e.g. for production of transgenic rape plants with
CC low or modified oil content, diversion of metabolism to alternative
CC storage compounds, e.g. starch, protein or engineered polymers, or
CC production of plants with enhanced oil content. The DNA may also
CC be used as a probe to obtain similar genes from other plants. The
CC transit peptide may be used to direct other proteins to leaf

CC plastids.
XX SQ Sequence 315 AA;

Query Match 9.78; Score 206.5; DB 17; Length 315;
Best Local Similarity 27.48; Pred. No. 1e-11;
Matches 69; Conservative 45; Mismatches 95; Indels 43; Gaps 7;

```

OY      13 VETIGASRGICKAVIALNKAQGANIVI---AKTAOPHPRLGTITTAABEIAVAGKA 68
Db      75 vvtvgasrglqakalalslgkagckvlnyarsakee-----evskqlaeaygqa 124
OY      69 LPCIVDVDEQOISAAYEAKAIKKEGIDILVNNSAISLTNTLTPPKRLDLMNVNTRG 128
Db      125 ltfgdvskaeadeaammktaldawgtldvvnngitrdtlllmkksqdwedvldnltg 184
OY      129 TYLSKACIPILKSKSVAAHILNISPLNLNVKRONCAVITIAKYSMTYLGMAEEFKG 188
Db      185 vflctgaatkimkkkrxgrlntasvyl--lgnlganyaaakag---vlgfkskaar 238
OY      189 E-----IAVNALMP-----KTAIHTAAMDMLGSGPIESQCRKVDIADA 227
Db      239 egasrnlnvnnvcpvgfflasdmntaklgedmekkllgtlplgrgqp--edvavglvefials 296
OY      228 AYSIFQPKPSFT 239
Db      297 paasyltgqft 308

```

RESULT 10

R89322
ID R89322 standard; Protein; 315 AA.

AC R89322;

DT 13-APR-1996 (first entry)

DE Rape seed beta-ketoacyl-ACP-ketoreductase.

KW Rape; seed; beta-ketoacyl-ACP-ketoreductase; beta-ketoreductase;
KM plasmid pJR10.1; CDNA library; embryo; Escherichia coli; vector;
KW plasmid; stroma; transit peptide; cassette; antisense; oilseed;
KM transgenic plant; crop improvement; lipid; metabolic engineering;
KM polymer; rapeseed oil.

OS Brassica napus.

FH Key Location/Qualifiers

FT Peptide 1..56 /note= "transit peptide"

PN W09602652-A2.

PD 01-FEB-1996.

PF 17-JUL-1995; 95WO-G801678.

PR 20-JUL-1994; 94GB-0014622.

PA (ZENE) ZENECA LTD.

PI Chase D, Elborough K, Fentem PA, Slabas AR, White A;

DR WPI: 1996-105914/11.

DR N-PSDB; Q99304.

PT New isolated rape beta-ketoreductase DNA - used to develop plants

PS Claim 1; Page 15; 29pp; English.

CC The sequence corresponds to a rape seed beta-ketoreductase encoded

CC by a cDNA insert in plasmid pRSL0.1 in *Escherichia coli* XL1-Blue.
CC
CC A plastid stroma targeting transit peptide is present. DNA
CC encoding the protein may be inserted in a vector or expression
CC cassette in sense or antisense orientation for expression in an
CC oilseed plant, e.g. for production of transgenic rape plants with
CC low or modified oil content, diversion of metabolism to alternative
CC storage compounds, e.g. starch, protein or engineered polymers, or
CC production of plants with enhanced oil content. The DNA may also
CC be used as a probe to obtain similar genes from other plants. The
CC transit peptide may be used to direct other proteins to seed
CC plastids.

Sequence 315 AA;

Query Match	9.7%	Score 206.5	DB 17	Length 315
Best Local Similarity	27.4%	Pred. No. 1e-11		
Matches 69	Conservative 45	Mismatches 95	Indels 43	Gaps 7

```

Oy 13 VEITGANSIGIKRATALKAKKAGANI---AAKTADPHRKLTGITYAAEEIEAAGKA 68
Db 75 vvvvgasrsgjkskatalalsgkagckrvllyharsaeae-----evakgleay99ga 124
Oy 69 LPCIYDVHDEEOISAAVEKAIKKEGGIDILVNBASAI$TLNTLDTPTKRILDLMMNVTRG 128
Db 125 itfgdvskeadveemmktaldaugtildvvvnnagiftrdliimkksqgdevdhlntg 184
Oy 129 TYL$KACIPIPLK$SKV$HIIINISPRILN$NVW$KQHCATIANKYG$SMVILV$AE$EFG 188
Db 185 vflcqaactlmmkkrkrglinlesvvgl--lgnlganyaa$aa$g---vlg$sk$aar 238
Oy 189 E---IAVNALMP-----KTAHT$TAM$MLGG$PGIESO$RKVDITADA 227
Db 239 egasrnlhvvvpcgfflasdm$taklgedmekkllgtlprlgr$ygp--edv$aglyefals 266
Oy 228 AYSIQKPKSF$T 239
Db 297 paasyltqgait 308

```

RESULT	ID		
B42640	ID	B42640 standard; protein; 313 AA.	
XX	AC	B42640;	
XX	DT	08-FEB-2001 (first entry)	
DE	XX	Human ORFX ORF204 polypeptide sequence SEQ ID NO:4808	

KM Human; open reading frame; ORF; detection; cytostatic; hepatotropic;
 KM vulnerability; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
 KM anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
 KM immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KM hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KM antiviral; antibacterial; antifungal; antihemmatic; antihypert;
 KM antinaeemic; gene therapy; cancer; proliferative disorder; hypertension;
 KM neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KM cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KM cholesterol ester storage; systemic lupus erythematosus; infection;
 KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KM bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KM thrombosis; contraceptive.

OS Homo sapiens.

PN W0200058473-A2.

PD 05-OCT-2000.

PF 31-MAR-2000; 2000WO-US08621.

XX

PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.

PA (CURA-) CURAGEN CORP.

PI Shimkets RA, Leach M;

DR WPI; 2000-602362/57.

DR N-PSDB; C76849.

PT	Novel nucleic acids and peptides derived from open reading frame X
PT	useful for treating e.g. cancers, proliferative disorders,
PT	neurodegenerative disorders and cardiovascular disease -

PS Claim 11; Page 3990; 5507pp; English.

CC C74446 C77606 encode the proteins given in B40237 to B43937, which
CC sequences the human OREX open reading frames 1 to 3161. The OREX
CC sequences have activities such as: cytostatic; hepatotropic; vlnetraty;
CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antihematic; antithyroid; and antinaemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC OREX-associated disorder. The nucleic acids can be used to express OREX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease, to
CC enhance coagulation; to inhibit thrombolysis; and as a contraceptive.

50 Sequence 313 AA;

Query Match	9.5%;	Score 203.5;	DB 21;	Length 313;
Best Local Similarity	25.4%;	Pred. No. 1.9e-11;		
Matches 72; Conservative	54;	Mismatches 118;	Indels 39;	Gaps 12

Qy	10	GCIVRTASGICAKILMKAOKGANVIAKTAOPKRLGITYTAAEEIYAVGKAL	69
Db	7	gqvcvvtgajjlgjlgjlgjlgckagatyltgr-----h--ldtlrvagaag199qcv	59
Qy	70	PCIVDVREBOQISAAVEAFAIK-KFGGIDILVNNSA--ISLTNT-----LDPFKRIDLM	121
Db	60	pvcvdsqesqevsrifegvdreqgrldvlvlnayagqtlitfrnkafwelpasmwdl	119
Qy	122	MAVNTRGTYLASKACIPILKSKSAVHLINISPLMLNPVWFKOKCATYIAKYGSMVLG	181
Db	120	mvgvltlghyfcsvygarlmvpagqgljlvjsspslg-----ymfnpvygvaacdrlaad	176
Qy	182	MAEEFKGE-IAVNALMP---KTAI---HTAAMDIMGGGIE-----SCKRKYDIIADAA	228
Db	177	caheirldrgvscvslwpljvgltellkemaakeavlgqrvljqkfssafssaetelsgcv	236
Qy	229	YSIFOKPK--SFTGNFVIDENLIKKEEGLENDVYAIKRGHLO	269
Db	237	valatcpmlisgkvlpjpscdarlgyrltdv-----grfvq	273

RESULT 12

ID Y27004 standard; Protein; 313 AA.

AC Y27004;

XX

DT	20-SEP-1999	(first entry)
XX		
DE	Human short-chain dehydrogenase (HSCD) enzyme.	
XX		
KW	Short-chain dehydrogenase; HSCD; enzyme; pyruvate; coenzyme A; human;	
KM	mitochondria; immune disorder; cancer; leukemia; adenocarcinoma;	
KW	lymphoma; breast; lung; testis; prostate; brain; Addison's disease;	
KM	acquired immune deficiency syndrome; asthma; anemia; Crohn's disease;	
KW	Graves disease; AIDS; gene therapy.	
XX		
OS	Homo sapiens.	
XX		
PN	US5928923-A.	
XX		
PD	27-JUL-1999.	
XX		
PF	05-FEB-1998; 98US-0019216.	
XX		
PR	05-FEB-1998; 98US-0019216.	
XX		
PA	(INCY-) INCYTE PHARM INC.	
XX		
PI	Corley NC, Lal P;	
XX		
DR	WPI: 1999-429503/36.	
XX		
DR	N-PSDB: X86772.	
XX		
PT	Nucleic acids encoding human short chain dehydrogenase enzymes	
PT	useful for the diagnosis, prevention and treatment of disorders	
PT	associated with abnormal expression of the enzyme, such as immune	
XX	disorders and cancer	
XX		
PS	Claim 1; Fig 1A-D; 27pp; English.	
XX		
CC	This represents a human short-chain dehydrogenase (HSCD) enzyme. The	
CC	dehydrogenase enzymes catalyze an irreversible reaction between pyruvate	
CC	and coenzyme A, to form CO2 and the intermediate COA, in mitochondria.	
CC	Host cells containing vectors comprising the HSCD nucleic acid may be	
CC	used to produce the HSCD enzyme, according to standard recombinant DNA	
CC	methodology. The enzyme may then be used as an antigen in the production	
CC	of antibodies or in assays to identify antagonists of HSCD activity.	
CC	These antagonists may then be used to treat disorders associated with	
CC	inappropriate expression, or over activity of HSCD such as immune	
CC	disorders and cancers (the antagonist interferes with the reaction	
CC	between pyruvate and coenzyme A). For example, the antagonists may be	
CC	used to treat leukemia, lymphomas, adenocarcinomas and cancers of the	
CC	breast, lung, testis, prostate and brain, Addison's disease, acquired	
CC	immune deficiency syndrome (AIDS), asthma, anemia, Crohn's disease and	
CC	Graves disease. The nucleic acids and antisense sequences can be used	
CC	in gene therapy.	
XX		
SO	Sequence 313 AA:	
Query Match	9.5%; Score 202.5; DB 20; Length 313;	
Best Local Similarity	25.4%; Pred. No. 2.4e-11;	
Matches	72; Conservative 53; Mismatches 119; Indels 39; Gaps 12.	
QY	10 GCTVFTGASRCIGAKALILKAADGANIVIAAKTAOPHPKLLGTLYTAAEIEAVGKAL 69	
DB	7 GQVGVVLTGASRGISGALIGLQJCKAQATYITGR---h--Idtlrvvaqeeqslggcv 59	
QY	70 PCIVYVREQGISAAVEAKA-KKSGDILVNNASA--ISLNT-----LDTPTRLDLM 121	
DB	60 pvcvdsaseeserrllfegvdrtegggrldvlnmayvqtlntlnkafwelpasmwddl 119	
QY	122 MNVNTRGTYLASKACIPYLKSKSVAHILINISPLLNPVWEKHCACAYTIATKGSMTYVG 181	
DB	120 navgrltgyfscvygarilmvpeggllivlvspsglq--ymfnvpygygkaacklaad 176	
QY	182 MAEEKGE-IAVNALWP---KTAI---HTAAMDILGGEIE-----SOCRAVDIITADA 228	
DB	177 cahrtrtgvsvcsvlmpivgtellkellmakceevlqdpvlkfkfaissaeetlsgkcv 236	

```

Oy 229 KSIFOKPK--SPTGFNEVIDENTLKEEGEENFDPVIAIKGHPQ 269
      :: | | | : | : | : | : | : | : | : | : | : | : |
Db 237 valatcpnlislsqkvi pscdiarrylrdv-----grpvq 273

RESULT 13
ID R27757
AC R27757 standard; protein; 261 AA.
XX
XX R27757;
XX
XX 11-MAR-1993 (first entry)
XX
XX Glucose dehydrogenase.
XX
XX GDH; mutant; recombinant; mass production; tetramer; thermostable.
XX
XX Bacillus megaterium.
XX
XX Key Location/Qualifiers
FT Misc-difference 96 /note= "site of Glu->Val or Lys mutation"
FT FT Misc-difference 252
FT FT Misc-difference 252 /note= "site of Glu->Leu mutation"
FT FT Misc-difference 253 /note= "site of Tyr->Glu mutation"
FT FT Misc-difference 22
FT FT Misc-difference 43 /label= Ser, Ala
FT FT Misc-difference 43 /label= Asp, Glu
FT FT Misc-difference 79
FT FT Misc-difference 95 /label= Ala, Ser
FT FT Misc-difference 95 /label= Leu, Met
XX
XX JP04258293-A.
XX
XX 14-SEP-1992.
XX
XX 13-FEB-1991; 91Jp-0106927.
XX
XX 13-FEB-1991; 91Jp-0106927.
XX
XX (AMAN ) AMANO PHARM KK.
XX
XX WPI: 1992-354684/43.
DR
XX Glucose dehydrogenase prepn. using transformed recombinant DNA
PT from Bacillus megaterium - has specified transformations giving
PT glucose dehydrogenase-expressing vector, introduced into E.coli.
PT for culture
XX
XX Claim 1; Page 8; 12pp; Japanese.
PS
XX The glucose dehydrogenase enzyme is encoded by a recombinant DNA
CC clone from Bacillus megaterium. The DNA sequence may be mutated by
CC site directed mutagenesis to introduce mutations to the protein
CC sequence (see feature table). The DNA may be used to transform E.
CC coli cells, and transformants may be cultured to mass produce GDH.
CC The mutant GDH is not influenced by ion strength, exists as a
CC tetramer and is thermostable.
XX
XX Sequence 261 AA;

Query Match 9.3%; Score 199.5; DB 13; Length 261;
Best Local Similarity 29.1%; Pred. No. 3.6e-11;
Matches 58; Conservative 44; Mismatches 84; Indels 13; Gaps 6;

Oy 8 LAGCIVFTTGASRGIGKAIALKAADGANNIYAIAKTAQPHFKLLCTITTAEEITAVGCK 67
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5 legkavvtlgtstgkymalrfatekakvvnvyskexaan-----svleelkkvyge 58

```

```

0y      68 ALPCIVDRDQOISAAYEKAKRFGGIDILVNNASAIISLTNTDPTRKRLDIMMNVNTR 127
          ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      59 alavkgdtvesdvlnlysqikexfgkldvmlnnaagxenvpsvhemsldwnkvicnl 118
          ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0y      128 GTYLASKACIPYELKRSKV-AHILNISPLNLNPWFPHOCAYTITAKYGSMNYVLGNAEFP 186
          ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      119 gafisgrealkyfvendilqgvtlmsvshekfpplrvh--yaaskgsmllmetclaley 176
          ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0y      187 --KGELAVNALMPKTAIHT 203
          ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      177 apkq-irvnniqp-gaint 193
          ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT   14
R03846    R03846 standard; protein; 261 AA.
XX
AC        R03846;
XX
DT        09-AUG-1990 (first entry)
XX
DE        Modified glucose dehydrogenase gene of Bacillus megaterium.
XX
KW        Glucose dehydrogenase; Bacillus megaterium; heat stability;
RW        glucose assay; ss.
XX
OS        Bacillus megaterium.
XX
PN        DE3931716-A.
XX
PD        29-SEP-1988.
XX
PE        22-SEP-1989; 89DE-3931716.
XX
PR        22-SEP-1988; 88JP-0237699.
XX
PA        (AMANO ) AMANO PHARM KK.
XX
PI        Makino YS, Negoro S, Urabe IA, Okada HT.
XX
DR        WPI; 1990-100507/14.
XX
PT        New DNA encoding modified forms of glucose dehydrogenase -
PR        from Bacillus megabacterium, having specific amino acid
PS        replacements, with higher heat stability.
XX
PP        Claim 1; Page 13; 16pp; German.
XX
XX
XX        The normal glu-96 residue is replaced by Lys, Gly or Ala, Gln-252 by
CC        Leu or Tyr-253 by Cys. Ser-22 residue can also be Ala, Asp-43 can be
CC        Glu, Ala-79 can be Ser and Leu-95 can be Met.E.coli containing this
CC        sequence will produce a form of GDH with improved heat stability. These
CC        enzymes are useful for glucose assay and can be produced at low costs
CC        or in large amounts.
CC        See also Q03772.
XX
XX
SO        Sequence         261 AA;

Query Match           9.3%; Score 198.5; DB 11; Length 261;
Best Local Similarity 29.6%; Pred. No.4,5e-11;
Matches 59; Conservative 45; Mismatches 82; Indels 13; Gaps

0y      8 LAGCIVFTTGASRSGIKRAIALKAKADGANIYIAKTQAPHPKLIGTTYTAEELEANGVK 67
          ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      5 legkvvviltsgstglgskmaritratekxkvvnryrskdean-----svleeikvyge 58
          ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

0y      68 ALPCIVDRDQOISAAYEKAKRFGGIDILVNNASAIISLTNTDPTRKRLDIMMNVNTR 127
          ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      59 alavkgdtvesdvlnlysqikexfgkldvmlnnaaglxnpvsshemslsdwnkvicdnlw 118
          ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

0y      128 GTYLASKACIPYELKRSKV-AHILNISPLNLNPWFPHOCAYTITAKYGSMNYVLGNAEFP 186
          ||| | : : : : ~~~~~~
Db      119 gafisgrealkyfvendilqgvtlmsvshekfpplrvh--yaaskgsmllmetclaley 176
          ||| | : : : : ~~~~~~

0y      187 --KGELAVNALMPKTAIHT 203
          ||| | : : : : ~~~~~~
Db      177 apkq-irvnniqp-gaint 193
          ||| | : : : : ~~~~~~

```

Db	119	gafigsrtaikyfvendtkyfvlimssvhektpwlfh--yaaaskgmkmlmwexaleley	176
Qy	187	--KGEIAVNALMPKTAIHT	203
Db	177	apkg-irvnmigp-gaint	193
	RESULT	15	
	Y55986		
ID	Y55986	standard; Protein; 251 AA.	
AC	Y55986;		
XX	15-MAR-2000	(first entry)	
DT			
XX			
DE	5.clavuligerus cladh protein.		
XX			
XX	Operon: cladh; clar: biosynthetic pathway; clavulanic acid; enzyme;		
KW	clavulanate-9-aldehyde reductase; transcriptional regulator;		
KM	gene expression.		
XX			
OS	Streptomycetes clavuligerus.		
XX			
PN	ES2131001-A1.		
XX			
PD	01-JUL-1999.		
XX			
PF	16-JUN-1997; 97ES-0001305.		
XX			
PR	16-JUN-1997; 97ES-0001305.		
XX			
PA	(ANTI) ANTIBIOTICS SAM.		
XX			
DR	WPI; 2000-026016/03.		
XX			
DR	N-PSDB; Z30700.		
XX			
PT	Improving production of clavulanic acid by Streptomycetes - by		
PT	super-expression of clar gene.		
XX			
PS	Example 1; Page 8-9; 23pp; Spanish.		
XX			
XX			
CC	This sequence represents the cladh protein which is a putative		
CC	clavulanate-9-aldehyde reductase used in the biosynthetic pathway for		
CC	clavulanic acid in Streptomycetes clavuligerus. The invention relates to		
CC	methods of improving production of clavulanic acid by Streptomycetes		
CC	super-expression of clar gene. The clar gene is characterised in that		
CC	it is localised in the gene group encoding genes for biosynthesis of		
CC	clavulanic acid.		
XX			
XX			
XX	Sequence	251 AA;	
XX			
Qy	Query Match	9.3%; Score 198; DB 21; Length 251;	
Db	Best Local Similarity	27.8%; Pred. No. 4.7e-11;	
	Matches	69; Conservative 40; Mismatches 103; Indels 36; Gaps	
Qy	1	MLPNTGRAGCGVFPTGTASRGIRKAIKAAKDGANIVIAKTAQPHPKLGTITYAAEE	60
Db	4	mmpss--lqgkvaalitgrelghmrataralapagaavaiaarve-----klralgde	54
Qy	61	IEAVGKALPCIVDRDQOISAAVEKAIRKFGGIDIIIVNNAISAISLTNTLDTPTKRIDL	120
Db	55	ltaagakrhvleldvdrdgydaavaestealagldilvnmagimlqprvedadctdtr	114
Qy	121	MMNVNTGCTYASKACIPYLLKSK--VAHILNISPPMLNLPYWFQHCAYTIATKGMSMY	178
Db	115	midctnllglimymtraaalphllrskgtlvqgmssia-----grvtvrnaavygatkfgvna	169
Qy	179	VLGMAEEF-KGEIAVNALMPKTA-----IHTAAMDMLGSPGIE---SQCRLY---DI	223
Db	170	setvqetelergrvrvvleqptldetelghlthataikem-----yegrlsqrltklqagdl	224

Oy 224 IADAVSI 231
| : :
Db 225 aeavryav 232

Search completed: June 11, 2001, 16:23:10
Job time: 140 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2001, 16:23:15 ; Search time 29.17 Seconds
(Without alignments)
1679.565 Million cell updates/sec

Title: US-09-464-039-7
Perfect score: 2139
Sequence: 1 MHPNTRGLAGCTVFITGASR.....GNMALAIKIKELKIMQNMARL 418

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :
1: SPREMBL.15:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.rodent:*
13: sp.unclassified:*
14: sp.vertebrate:*
15: sp.virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the total score distribution, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1179.5	55.1	412	5	Q9VB10	Q9VB10 drosophila
2	1034	48.3	938	5	Q09979	Q09979 caenorhabdi
3	891.5	41.7	293	5	Q18639	Q18639 caenorhabdi
4	665	31.1	282	2	Q05842	Q05842 mycobacteri
5	351	16.4	441	5	Q9NKM1	Q9NKM1 dictyosteli
6	302.5	14.1	1015	3	Q9UVH9	Q9UVH9 glomus moss
7	286	13.4	436	5	Q21481	Q21481 caenorhabdi
8	267	12.5	248	2	Q44326	Q44326 agrobacteri
9	245.5	11.5	735	13	Q42484	Q42484 gallus gall
10	244	11.0	240	1	Q9UY54	Q9UY54 pyrococcus
11	235	10.9	246	2	Q9KA03	Q9KA03 bacillus ha
12	233.5	10.6	255	2	P95286	P95286 mycobacteri
13	227	10.6	253	2	Q9K377	Q9K377 streptomyce
14	227	10.6	679	2	Q9XAB2	Q9XAB2 streptomyce
15	224.5	10.5	301	2	Q53398	Q53398 mycobacteri
16	224	10.3	247	2	Q9LGV7	Q9LGV7 streptomyce
17	221	10.2	248	2	P71824	P71824 mycobacteri
18	218.5	10.2	251	5	Q9RH22	Q9RH22 zymomonas m
19	218	10.2	251	5	Q9U112	Q9U112 drosophila

20	217	10.1	256	2	P95273	P95273 mycobacteri
21	216.5	10.1	255	1	Q34187	Q34187 halobacteri
22	216.5	10.1	263	2	Q9RK87	Q9RK87 streptomyce
23	216	10.1	275	2	Q9ZB95	Q9ZB95 rhodococcus
24	215.5	10.1	263	2	Q07882	Q07882 staphylococ
25	214.5	10.0	246	2	Q9K636	Q9K636 bacillus ha
26	214.5	10.0	256	2	Q9R878	Q9R878 klebsiella
27	214.5	10.0	284	5	Q23618	Q23618 caenorhabdi
28	214	10.0	255	2	Q9L2C9	Q9L2C9 streptomyce
29	213	10.0	269	2	Q9L0G1	Q9L0G1 streptomyce
30	211.5	9.9	270	5	Q9W5C0	Q9W5C0 drosophila
31	211	9.9	279	5	Q16969	Q16969 caenorhabdi
32	210.5	9.8	285	2	Q9X4W7	Q9X4W7 pseudomonas
33	210.5	9.8	598	5	Q9VX10	Q9VX10 drosophila
34	208	9.7	265	2	Q9S2E4	Q9S2E4 streptomyce
35	208	9.7	319	5	Q9N538	Q9N538 caenorhabdi
36	206.5	9.7	257	5	Q9VNF3	Q9VNF3 drosophila
37	206.5	9.7	257	11	P70540	P70540 ratius norv
38	206	9.6	250	2	Q56840	Q56840 xanthobacte
39	205.5	9.6	255	2	Q9PC02	Q9PC02 xyella fas
40	205	9.6	266	2	Q69366	Q69366 rhodococcus
41	204	9.5	249	2	Q9L8G5	Q9L8G5 bacillus th
42	202.5	9.5	254	5	Q9N350	Q9N350 caenorhabdi
43	202.5	9.5	1639	5	Q9VNF2	Q9VNF2 drosophila
44	201.5	9.4	734	11	P70523	P70523 ratius norv
45	200.5	9.4	281	2	Q9K5C3	Q9K5C3 burkholderi

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	412 AA.
Q9VB10	Q9VB10	Q9VB10		
AC	Q9VB10:			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)			
DE	CG5590 PROTEIN.			
GN	CG5590.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_Taxid=1227;			
RP	[1]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=BERKELEY;			
RX	MEDLINE=20196006; PubMed=10731132;			
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
RA	Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,			
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,			
RA	Abul J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,			
RA	Ballev R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,			
RA	Besoon K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintsov S.,			
RA	Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,			
RA	Buttis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,			
RA	de Palos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,			
RA	Dubin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,			
RA	Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,			
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
RA	Harris N.L., Harvey D., Heilmann T.J., Hernandez J.R., Houck J.,			
RA	Hoskins D., Houston K.A., Howland T.J., Wei M.-H., Iyegbem C.,			
RA	Jatall M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,			
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,			
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,			
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,			
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,			

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Piltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relbert K., Remington K., Saunders R.D.C., Scheller F., Shen H.,
 RA Shue B.C., Sidenkimos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.W., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Glubs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003762; AAF56735.1; -
 DR HSSP: P47227; 1BD8.
 DR FLYBASE: FBgn0039537; CG5590.
 DR INTERPRO: IPR002198; -
 DR INTERPRO: IPR002347; -
 DR INTERPRO: IPR003033; -
 DR PFAM: PF00106; adh_short; 1.
 DR PFAM: PF02036; SCP2; 1.
 DR PRINTS: PRO0081; GDHRDH.
 SO SEQUENCE 412 AA; 44354 MW; 21365F07FCAFE0B CRC64;

Query Match 55.1%; Score 1179.5; DB 5; Length 412;
 Best Local Similarity 56.4%; Pred. No. 3.9e-75;
 Matches 238; Conservative 62; Mismatches 103; Indels 19; Gaps 5;

QY 4 MNGRLAGCVFTGASRGIGKAIKAAKADGANIYAATAOPHRLTGTIYAAEIEA 63
 DB 3 MNGRLAGRLFTGTASRGIGKAIKAAKADGANIYAATAOPHRLTGTIYAAEIEA 62
 QY 64 VGGKALPCIVDRDEQOISAAVEKAIRKFGIDILVNNASISLTNTLDTPTKRLDMMN 123
 DB 63 AGCKAPCVDRDEQOISAAVEKAIRKFGIDILVNNASISLTNTPTDMKRYDLMN 122
 QY 124 VNTRGTYLAKACIPYLKSKVAHILNISPLNLPWFKQHCATIIAYGSMYVLGMA 183
 DB 123 INTRGFELVSKCLPYLKSNAHILNISPLSMKMGFPYHAYGMAYGSMCYLMA 182
 QY 184 EEFGKE-IVNMLPMTAHTAAMDLMGPGIESOCRKDIADAYSTF-OKPKSFTGN 241
 DB 183 AERKDEGISVNLMPRTAHTAEMLTGPDSAKSKRPEIMADAYALTREPROSTQ 242
 QY 242 FVIDENILKEEGIEENFYAI--KPGHPLOPDFLDEYPEAVSKKVESTGAVPEFEKEL 299
 DB 243 FVYDDELVSAGITDLETAACFRENADKLMPDFVEEKGAPVENEAAADDA----- 294
 QY 300 QIQPKRSGA---VEETFRIVKDSLDVYKATQATLYLFELSGEDGCTPFLDKSGGAV 356
 DB 295 ---AASGSDVKIQPLFKRIEISLSPETIVSKTQAFQFNISGAEQGTWFLDKNGSGC 350
 QY 357 GCGEPDQADVYMSMTDDFVFMFSGKLPTMAFMSGKLTKIGNMALIKLEKLNQMA 416
 DB 351 GAGTPAAPDALTINMSKNFPMFSGKLAAAPYMTGKLISDFOKALKLEKLMALKS 410
 QY 417 RL 418
 DB 411 KL 412

RESULT 2
 ID 009979 PRELIMINARY: PRT: 938 AA.
 AC 009979:
 DT 01-JUN-1998 (TREMblrel. 06, Created)
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
 DE HYPOTHETICAL 105.9 KDA PROTEIN C17G10.8 IN CHROMOSOME II.
 GN C17G10.8.
 OS *Caenorhabditis elegans*.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; *Caenorhabditis*.
 NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Johnson D.;
 RL Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (DEC-1997) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: STRONG IN THE C-TERMINUS TO C.ELEGANS C45B11.3 AND
 CC MYCOBACTERIUM TUBERCULOSIS MTCY07D11.02C.
 DR EMBL: U28739; AAB93456.1; -
 DR MORMEP: C17G10.8; CE16861.
 DR INTERPRO: IPR002198; -
 DR INTERPRO: IPR003033; -
 DR PFAM: PF00106; adh_short; 1.
 DR PFAM: PF02036; SCP2; 1.
 KW Hypothetical protein.
 SO SEQUENCE 938 AA; 105896 MW; D64A09FBBC76C775 CRC64;

Query Match 48.3%; Score 1034; DB 5; Length 938;
 Best Local Similarity 54.6%; Pred. No. 2.4e-64;
 Matches 230; Conservative 52; Mismatches 111; Indels 28; Gaps 8;

QY 7 RLACGCVFTGASRGIGKAIKAAKADGANIYAATAOPHRLTGTIYAAEIEAVGC 66
 DB 526 KEVGRVLTGTASRGIGKAIKAAKADGANIYAATAOPHRLTGTIYAAEIEKAGC 585
 QY 67 KALPCIVDRDEQOISAAVEKAIRKFGIDILVNNASISLTNTLDTPTKRLDMMNVT 126
 DB 586 KALPCIVDRDEASVKAIVEAKRFGIDILVNNASISLTNTTERTEMKRYDLMSINT 645
 QY 127 RGTLYLAKACIPYLKSKVAHILNISPLNLPWFKQHCATIIAYGSMYVLGMAEF 186
 DB 646 RGTFLMTKCLPYLKSNAHILNISPLMETRMFANHVATMAKYGSMCYLMAEF 705
 QY 187 KGE-IVNMLPMTAHTAAMDLMGPGIESOCRKDIADAYSTFQK-PRKSTGNFY 244
 DB 706 RRGIVNMLMPRTAHTAEMLTGPDSAKSKRPEIMADAYALTREPROSTQ 765
 QY 245 DENILKEEGIEENFYAIKPGHPLOPDFLDEYPEAVSKKVESTGAVPEFEKELQDPK 304
 DB 766 DEDILKAEVTDPRVACVPDAPLMPDFI---PACTYDHKFSGA-----QIGKK 813
 QY 305 PR---SGAVEE---TFRIYKDSLDVYKATQATLYLFELSGEDGT-----WFLDKSK 352
 DB 814 NKTHEGAVVEEIKQIFTSAKRLNLNDIVKKTGEVYFLL--KDPTTKSERITITDLKNG 871
 QY 353 GGNVGEPSDQADVYMSMTDDFVFMFSGKLPTMAFMSGKLTKIGNMALIKLEKLN 412
 DB 872 EGALTPKKAAGRADVFTLAPEHAPLFTGKLRLPTALTMTKLIQISGDPMGAKLESLLR 931
 QY 413 Q 413
 DB 932 K 932

RESULT 3
 ID 018639 PRELIMINARY: PRT: 293 AA.
 AC 018639:
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE C45B11.3 PROTEIN.
 GN C45B11.3.
 OS *Caenorhabditis elegans*.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

CC Rhabdilitidae: Peloderinae: Caenorhabdilis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnson L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Koopa A., Saunders D., Sounkseen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierley-Mieg J., Thomas K., Vaund M., Vaughan K., Waterston R.,
RA Watson A., Wellstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans".
RL Nature 368:32-38(1994).
DR EMBL: Z74029; CAAS8431.1; -.
DR INTERPRO: IPR002198; -.
DR PRAM: PF00106; adh_short; 1.
SQ SEQUENCE 293 AA; 31524 MW; 7277E0164310AA2B CRC64;

Query Match	41.7%	Score 891.5	DB 5	Length 293
Best Local Similarity	61.7%	Pred. No. 4.1e-55		
Matches 185	Conservative 29	Mismatches 73	Indels 13	Gaps 3

```
QY      2 LPNNGRSLAGCGVFTTGGASRGIGKIALTKAANDGNNIYYAKTAQDPHRKLGTITYAAEI    61
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      5 IKNTGFPAKGKVFTTGASRGIGKEIALTKAKDGNIIYAAKTATAHPRKLGCTITYAAEI    64
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      62 EAVGGAALPCIVDRDEQQISAAVEKAIRKKGIDILVNNAISITLNTLDTPTRKIDM    121
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      65 EKAGAGHMLPCVVADVRDEDAVAAYADAARFKFGIDIIINNASASTLNTEDTDMKRXYDM    124
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      122 MNVNVRTGYTLASKRACIPYLKKSKVAHLNTSPRLNLVPVNEKHCAVTIAIKYGSMTVLG    181
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      125 HSINTRGTYLLTKTCLEPYLRKKGKPHVLTNSPRLDMDAKKNGFHPVGVTYTKMKFGMSKCVLG    184
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      182 MAEEFKG - ELAVNALMWERTAIHTAAMDMLGGPGIESOCRRVDIITADAAYSIFOK - PKSFT    239
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      185 HHEEFRRPGIIVNALWLPALTATWTSAAMEFLSHSGDSACNRKASIMWADSAYAILSDSKKPT    244
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      240 GNFIPEIDILLKEBGIENFDVYAIRGPHLPDPLDLEYPEAVSKSVYSTGAVPEPKREKL    299
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      245 GNFILDEDLLOGITITNEEOIECVPGSSLLPDFFVPKRSFA-----PFEGSSKL    293
```

RESULT	4			
005842		PRELIMINARY;	PRT;	282 AA.
ID	005842			
AC	005842;			
DT	01-JUL-1997 (TREMblrel. 04, Created)			
DT	01-JUL-1997 (TREMblrel. 04, Last sequence update)			
DT	01-JUN-2000 (TREMblrel. 14, Last annotation update)			
DE	HYPOTHEtical 29.8 kDa PROTEIN.			
GN	RV3224 OR MTCY07D11.02C.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID-1773;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-H37Rv;			
RX	MEDLINE-98295987; Pubmed-9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			

RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Honrsny T., Jaegels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Ratandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.:
RT "Deciphering the Biology of *Mycobacterium tuberculosis* from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
DR EMBL: Z95120, CAB08313.1, -.
DR TUBERCULIST: RV3224; -.
DR INTERPRO: IPR002198; -.
DR PFAM: PFO0106; adh_short; 1.
KW Hypothetical protein.
SQ SEQUENCE 282 AA: 29814 MF: C1A3555EC2E24432 CRC64;

Query Match	31.18	Score 665	DB 2	Length 282
Best Local Similarity	48.98	Pred. NO. 3.3e-39		
Matches 133	Conservative 57	Mismatches 78	Indels 4	Gaps 4

QY 8 LAGCTVFTTGSRRIGKATLAKKANDKANVIYAKTPRKLILGTYTAAETEAIVGK 67

Db 3 LNGTHMFTISGSRIGLAIARARADCANLTALIKTAEPRKLPGLTYFTAAKLEEBNGGQ 62

QY 68 ALPCIVDVRDEQOISAAVEKAIKKEGGIDILVNNASISLTNTLDTPYTKRLDLMNVNTR 127

Db 63 ALPTVGDIRDPDAAVASAVATVEQFSGIDICVNNASAINIGSITREYMKRFEDLNGIOVR 122

QY 128 GTYLASKACIPYLKSKVAHILNTSPLNLINPVRKQHCAYTAKYMSYVYLGMAEFP 187

Db 123 GTYAVSQACIPYHMKGREPHILNTSPLTLEKKLVR-TAYMAKYQMTICALGIAIEEMR 181

QY 188 GE-TAYNALMFKTAIHTMA-DMLGSGIEQCKRVIDIDAAYSIFQPKRS-FTGNFVI 244

Db 182 ADGIASTNLMPRTWATAAAYONLIGSGDEAMARSRKPEYADAAYIVNKPATEYTKTLL 241

QY 245 DENILKEGIEIENPDVYAIKRGHP-LQDPDFIDE 276

Db 242 CEDVLVESGVDTLSVYDPCPATLGVGLWVED 273

RESULT	5		
Q9NKM1			
ID	Q9NKM1	PRELIMINARY;	PRT; 441 AA.
AC	Q9NKM1;		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)		
DE	HYDROXYSTEROID DEHYDROGENASE 4 HOMOLOG.		
GN	DDHSD4.		
OS	Dictyostelium discoideum (Slime mold).		
OC	Eukaryota; Dictyostelid; Dictyostelium.		
OX	NCBI_TaxID=44689;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=KAX4;		
RA	Maeda M., Kuwayama H.;		
RT	"Dictyostelium HSD4-homolog.";		
RL	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.		
SR	EMBL; AB042104; BAA94961.1; -.		
DS	SEQUENCE 441 AA; 46387 MW; 2CEE3D9779D93BCA CRC64;		

```
Query Match      16.4% Score 351; DB 5; Length 441;
Best Local Similarity 28.3%; Pred. No. 8.1e-17;
Matches 129; Conservative 69; Mismatches 172; Indels 86; Gaps 16.
```

QY 13 VFTGASGRICKAKALTKKAKDGANIVI-----AAKRAOPHPKLGTITVNA-----EEIEAV 64
| : ||| | ||| | : | : | : | : | : | : | : | : | : | : | : | : |
Db 10 VIIVGAGGICIKVVALEEFARKGAKRVVNDLGSHTGQ-----GSSSKRADVVEIKKA 63

QY 65 GKALPCTIVDVDEQQISAAYEAKIRFGGIDILVNNSAISLTNTLDTPPKRLDIMMNV 124
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

Db 64 GGAIVANYSDEGEK1---VQTAMDSFGVDILINNAGILRDVSGFKMTDGDMDLVYRV 120
QY 125 NTRGTYLASKACIPYLTKSKVAHLINISPLNLNPFVFKOHCATTIARYG-MSMYVLG-- 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 HAAGAVYLSAAMNHMEKNGFRIMTSSAAGL-----YGNQNGAANGSMKALGSL 173
QY 182 --MAEERKE-IAVNAIMPRTAHTTAAMDMLGPGIESQCRKVDIADAAYSTFOKRSF 238
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 174 NTLAQEGSKNHNHNTIAPITAA--SRLETESVMPPEILEOKMPDVIPLVLVYLCHQDTTET 231
QY 239 TGNF-----VIDENILKEE-----GIENFQVYAIKPGHPLPDPEFLD 275
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 232 GGFVFNAGVSKVRLOKRSAGVYWKDLTPREKIDNMAQIESFD---NPSTPISAS---- 283
QY 276 EYPEAVSKVESTGAVPEFEKEKLOLPK---PRSGAVEETFRIVKDSLS----- 322
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 284 ---ESVSGITLAIVNSKPADGESVLRPRKYAVPALAATPSGVSVDVGNASKIETTTIG 340
QY 323 -----DDVYKATQATILFEL-SEGDGTFPLDLKSKGNVGEPESDQADVYMSMTTD 375
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 341 NIGAKGAEIWKINGIYLINIKKGTNTQAWALDLKNGSGSIVGAGSTKPRVITVDEDED 400
QY 376 FVKMFSGLKPTMAFMSGKLIKGNMALAIKLEKLM 411
Db 401 FVDIMTKLNAQSAFTGKLIKISGMGLATKLGALM 436

RESULT 6
Q9UVH9 PRELIMINARY; PRT; 1015 AA.
AC Q9UVH9;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE FOX2 PROTEIN.
OS FOX2.
OS Glomus mosseseae.
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Glomales; Glomaceae;
OC Glomus.
OX NCBI_Taxid-27381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BEG 12;
RA Requena N., Petra F., Philipp F.;
RT "Molecular characterization of GmFOX2, an evolutionarily highly
   conserved gene from the mycorrhizal fungus Glomus mosseseae, down-
   regulated during interaction with rhizobacteria.";
RT Mol. Plant-Microbe Interact. 12:934-942(1999).
DL EMBL; AJ243538; CAB53552.1; -.
DR HSSP; Q12634; 1YBV.
DR INTERPRO; IPR002198; -.
DR INTERPRO; IPR002347; -.
DR INTERPRO; IPR002539; -.
DR INTERPRO; IPR003033; -.
DR PFAM; PF00106; adh_short; 2.
DR PFAM; PF01575; Maoc_dehydratas; 1.
DR PRINTS; PR00080; SDRFAMILY.
DR PRINTS; PR00081; GDHRH.
DR PROSITE; PS00061; ADH_SHORT; UNKNOWN_1.
SQ SEQUENCE 1015 AA; 108897 MW; FSA66FA50EAD51B-CRC64;

```

Query Match 14.18; Score 302.5; DB 3; Length 1015;
 Best Local Similarity 19.88; Pred. No. 7; Se-13;
 Matches 146; Conservative 66; Mismatches 179; Indels 345; Gaps 16;

```

QY 10 GCTVFTGASRGICAKALAKADGANIVIAKTAQPHKILGTITAA---EEIYAVG 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 274 GKAVIVGAGGGLGKRAVALLLGKLAGASVYVNDLGSVAHGQ--GATSSAAKQVVEIRAG 331
QY 66 GKALPCIVVDDEQGISAAVEKAIKRGGIDILVNNASAI SLNTLDTPTKRLDLMNVN 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 332 GKAVANYDSVEDGEK1---VETAIKAFGRVDIINNAGILRDKSEAFARMTDQMDLVORVH 388
QY 126 TRGYTLASKACIPYLTKSKVAHLINISPLNLNPFVFKOHCATTIARYGSMYVLGMAEE 185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 389 LRGYTKYTKAAMPYLTKOKGRITNTASSVGLGNFQGAN--YSTALGLIGFSNTIALE 446
QY 186 -----FKGEIYVNA-----LMPKTAHTTAAMDML----- 209
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 447 GRKNNIENVNTIAPNAGTRMTATIPMDMVEAFKPDVAVAPVSLAHBACPSTGNFEVGG 506
QY 210 -----GGPGI-----ESQCRKVDI-----A 225
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 507 GWAQVYQWRAGGIGFTPTSKALTPEDTASKIDITITNDDGRATHPTTQQAALQOFENFA 566
QY 226 DAAYSIFOKPKSFPTGNFVIDENILKEE-----GI-----E 255
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 567 NAQSESGQSKSKNSKNGKIDVEAKKRRFEPHYEYKERDVMYALGIGATRKLOQVYE 626
QY 256 NEDVYAIKPGHPLPDPEFL----- 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 627 NSDNFSVPIPTGVIPATILSNLTPLESEVLGDFNVMLLHGBOYLELKKPIPTSGKLITSP 686
QY 275 -----DEYPEAV----- 281
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 687 YVIDILDKGAVSFIFGITTTHDEKGEVIFENQTTLFTIRIGIGFGCKTGDDRCANATASNI 746
QY 282 ----- 281
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 747 PPKRAPDVYVKEKTENQNALYRLSGDYNPLHIDPSAMGCFPVPLHGMCTFGISGKH 806
QY 282 -----SKVESTGAVPEFEKEKLOLPK----- 302
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 807 IFSTFGKNDPTEFKSIRAKRLAVALPFPETLETQWKKDGDKVIEFQTRVERDVICIASAAV 866
QY 303 -----PKRSGAVEETFRIVKDSLS-----DDVYKATQATY 333
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 867 ELRGSASGASMGAPSAKTPSAGISVPGFOSSVFEQLKAGIDASPAREQAQVAKVGSF 926
QY 334 LFEISGDEG--GTFPLDLKSKGNVGEPESDQADVYMSMTTDDFVFMFSGLKLPYAFM 391
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 927 QLNKNAEKGESWVIDPKTGDGAVGIGPSKKADATIGVSDAPFMFLASGKLNAQAFM 986
QY 392 SGKLIKGNMALAIKL 407
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 987 SGKLIKGNMMLATKL 1002

RESULT 7
Q21481 PRELIMINARY; PRT; 436 AA.
AC Q21481;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE SIMILAR TO NON-SPECIFIC LIPID TRANSFER PROTEIN.
GN M03A8.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderae; Caenorhabditis.
OX NCBI_Taxid-6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-94150718; Pubmed-7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
   Jones M., Kershaw J., Kirsten J., Lalister N., Latreille P.,
   Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
   Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
   Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;

```



```
DB 2 LGKTAIVTGSARGIGRATAMELARGANVYVYVNGNKEKAERYVA-----EIKELGV 54
OY 67 KALPITVYRDEQOISAVERKAIRKFGIDILVNNASAI SLTNTLDTPTKRL-----DLA 121
DB 55 EAIATADVADESESOAVAKKETIDPFAGVDILVNNAGI-----TRDNLFMRKEEDMAY 109
OY 122 MNVTRGTYLASKACIPYLKSKVAHILNISPLNLNPVWFKOHCAVYTIKAGSMYVIG 181
DB 110 IDTNLKGFFHCKSKAVTRPMKORFRIINVSVC-----ALGNAGQANVYAA 157
OY 182 MAEEKGEIAVNAVMPKTAIHTAAMDMLGPGIESQCRKVDIADAAYSIFOKPRKFTGN 241
DB 158 KA-----GV 161
OY 242 FVIDENILKEGIEFNDVYAIKRGHPLOPDPFLDEYPEVNSKVESTGAVP-FKEEKLQ 300
DB 162 IGLVITLRELANRNIYVNAAPG-----FTMTGELPBDVNAQMLG 205
OY 301 LQPKRSGAVETFRIVKDSDDVVKAT-QAIVL 334
DB 206 QIPLARLGQPEEVAKAVRFLASDDASLYLGTIHY 240

RESULT 12
P95286 PRELIMINARY; PRT; 255 AA.
ID P95286;
AC P95286;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE HYPOTHEtical 27.0 KDA PROTEIN.
GN RV1928C OR MTCY09F9.36.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL; Z84498; CAB06498.1; -.
DR HSSP; P50163; ZAE1.
DR TUBERCULIST; RV1928C; -.
DR INTERPRO; IPR002198; -.
DR INTERPRO; IPR002347; -.
DR INTERPRO; IPR003015; -.
DR PFAM; PF00106; adh_short; 1.
DR PFAM; PF00678; adh_short_C2; 1.
DR PRINTS; PRO0080; SDRFAMILY.
DR PRINTS; PRO0081; GDRDH.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
DR PROSITE; PS00061; ADH_SHORT; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 255 AA; 27031 MM; 454692E6FAA253FF CRC64;
```

```
DB 9 LHGRALITGASTGIGKRVATLAVENAGVATARHDALETL-----ADEIGSGK 61
OY 68 ALPCTIVRDEQOISAVERKAIRKFGIDILVNNASAI SLTNTLDTPTKRLDLMNNVTR 127
DB 62 VPPCCDVSOHQOVTSMIDQVTAELGIDIAVCNAGIITVTPMLDMPLEEFORLONTWT 121
OY 128 GYLLASKACIPYLK-----SKVAHILNISPLNLNPVWFKOHCAVYTIKAGSMYVIG 176
DB 122 GVFLLQAQAANKVAKVQGGCVIINTASGHTIN-----PQOVSHYCA--SKAAVT 171
OY 177 MYVLMAGEEF-KGEIYVNAVMP 197
DB 172 HLTAMAVELAPHKIRVNSVP 193

RESULT 13
O9K3Y7 PRELIMINARY; PRT; 253 AA.
ID O9K3Y7;
AC O9K3Y7;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE PUTATIVE SHORT CHAIN OXIDOREDUCTASE.
GN 2SCG61.28C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Oliver K., Harris D.;
RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinash H., Hopwood D.A.;
RA "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-86(1996).
DR EMBL; AL359949; CAB95803.1; -.
SQ SEQUENCE 253 AA; 25937 MM; 608F0D5C1AE55A8A CRC64;
```

Query Match 10.6%; Score 227; DB 2; Length 253;
Best Local Similarity 31.1%; Pred. No. 1.8e-08;
Matches 64; Conservative 33; Mismatches 83; Indels 26; Gaps 5;

```
OY 15 ITGASRGIGKATLAKAAGANIVI-----AKRAQHPKLGITTYAAEIEVAGKA 68
DB 11 VTGARIGGATATLALAEGRVAVIIDEAACK-----DTEKITTAAGKA 57
OY 69 LPCIVDVREDOISAVERKAIRKFGIDILVNNASAI SLTNTLDTPTKRLDLMNNVTRG 128
DB 58 IAVGCDVSDQAQVAAVARIAEELGAPITLVNNAAGVLDNLLFKMSYSDMDITVNVHLRG 117
OY 129 TYLASKACIPYLKSKVAHILNISPLNLNPVWFKOHCAVYTIKAGSMYVIGMAEEF-K 187
DB 118 AFLMTKACQKHVDAKFGFRVNVLSSSSALGN--RGQVNSAKAKGLOGFTTAKELGK 174
OY 188 GEIYVNAVMP--KTAIHTAAMDVG 210
DB 175 FGVTAANAAPGFATYEMTKATADRVG 200

RESULT 14
O9XAB2
```

```
ID 09XAB2 PRELIMINARY; PRT; 679 AA.
AC 09XAB2;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE PUTATIVE OXIDOREDUCTASE.
GN SCFA3A.03.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K., Harris D.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb streptomycos coelicolor A3(2) chromosome.";
RN Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=843436;
RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
DR MBL; AL096837; CAB48890.1; -.
DR HSSP; P25529; IAHH.
DR INTERPRO; IPR002198; -.
DR INTERPRO; IPR002347; -.
DR PFAM; PF00106; adh_short; 1.
DR PRINTS; PRO0080; SDRFAMILY.
DR PRINTS; PRO0081; GDRHDH.
SQ SEQUENCE 679 AA; 71668 MW; 3826C2684B73BBA4 CRC64;
```

```
Query Match 10.6%; Score 227; DB 2; Length 679;
Best Local Similarity 29.4%; Pred. No. 8.3e-08;
Matches 82; Conservative 41; Mismatches 114; Indels 42; Gaps 11;
```

```
QY 2 LPNTGRLACCTVITGASRGIGKALAKAKDGANIYIAKTAOPHPKLLGTITYTAAEI 61
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 412 MPRPKPLATVALVTGSGSIGKAIARLVDGACVVAADLNAE-----NAAVA 461
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 62 EAVGG--KALPCIVDVNDEQOISAAVEKAIRKFGGIDILVNNASALSITNTLDTPTKRLD 119
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 462 EELGGDDKAAVAVVDVYSEEOIAAFQAAALAFGVDLVVNNAGISISKPLLETSAKDW 521
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 120 LMNVNVRGTLYLASKACIPYLKSKV-AHLINTSPRLNLPVWF-KQHCAYTTAKYMSM 177
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 522 LQHDIMARSGFVLSREARARVTAOELGCDIVYA---SKNAVFAGPNNIAYSATKADQAH 578
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 178 YVIGMAEEFGE--IAVNALMPTAHTAAMDMLGGPIESQCRKYDIIADAAYSIFQKP 235
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 579 QVRLAAEL-GEHGIRYNGVNPDSVVRSGI-FAGGWAAR-----AVYGV---P 624
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 236 KSFTGNFVIDENILKEGIEINFVYAIKPGHPLOPDFL 274
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 625 EEKIGEEFYAORTLLKRE-----VLPEHVANA VFAL 654
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 15
ID 053398 PRELIMINARY; PRT; 301 AA.
AC 053398;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
```

```
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE OXIDOREDUCTASE.
GN RV1050 OR MTV017.03.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jajels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence.";
DR Nature 393:537-544 (1998).
DR EMBL; AL021897; CAAL17166.1; -.
DR HSSP; P14061; 1FDU.
DR TUBERCULIST; RV1050; -.
DR INTERPRO; IPR002198; -.
DR PFAM; PF00106; adh_short; 1.
SQ SEQUENCE 301 AA; 32497 MW; 3B608922D291031C CRC64;
```

```
Query Match 10.5%; Score 224.5; DB 2; Length 301;
Best Local Similarity 30.9%; Pred. No. 3.5e-08;
Matches 58; Conservative 40; Mismatches 77; Indels 13; Gaps 5;
```

```
QY 13 VFTGASRGIGKALAKAKDGANIYIAKTAOPHPKLLGTITYTAAEIRAVGKALPCI 72
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 11 VLTGASSGIGENTAFKAFAREGAVVALAARE-----CALRRVAREIEAAGRAMVAP 63
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 73 VDVNDEQOISAAVEKAIRKFGGIDILVNNASALSITNTLDTPTKRLD--MNVNTRGTI 130
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 64 LDVSSSESVRAMVADVVEGGERIDVVFENNA-GVSLGPDVAETFLDDTREMELIDVLTGTV 122
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 131 LASKACIPYLKSKV-AHLINTSPRLNLPVWF-KQHCAYTTAKYMSYVIGMAEEFGE- 189
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 123 RVVREVLPIKKQORSGRIMNMSVYGRKA--FAFAGYSSAMHAIAGFSDALRQELRGSG 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 190 IAVNALMP 197
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 181 IAVSVIHP 188
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Search completed: June 11, 2001, 16:25:32
Job time: 137 sec
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 12, 2001, 11:40:23 ; Search time 1766 Seconds
(without alignments)
12668.826 Million cell updates/sec

Title: US-09-464-039-8
Perfect score: 2561
Sequence: 1 aggcagagaagtatgcaagca.....gsgmgaswmwawtwtammc 2561

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 9623517 seqs, 4368049070 residues

Word size : 0

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estp11:*
73: em_estp12:*
74: em_estp13:*
75: em_estp14:*
76: em_estp15:*
77: em_estp16:*
78: em_estp17:*
79: em_estp18:*
80: em_estp19:*
81: em_estp110:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	896	35.0	977	105	AL514158
2	802	31.3	861	106	AL528484
3	584	22.8	602	141	BE893726
4	500	19.5	887	168	BF699109
5	491	19.2	620	111	AM152065
6	489	19.1	963	141	BE892079
7	442	17.3	541	122	AW966207
8	419	16.4	487	5	AA315514
9	418	16.3	692	113	AW299683
10	398	15.3	829	171	BF979524
11	393	15.3	462	5	AA305116
12	330	12.9	484	157	W04752
13	326	12.7	482	156	R56767
14	325	12.7	386	9	AA622988
15	317	12.4	340	159	Z19446
16	296	11.6	305	6	AA348726
17	291	11.4	616	32	AV714646
18	284	11.1	497	32	AV750182

C	19	275	10.7	447	156	R89462	yq01b10. r1
	20	270	10.5	166	BE3924496	BE392446	601309277
	21	262	10.2	620	31	AV701344	AV701344
	22	262	10.2	661	31	AV701163	AV701163
	23	258	10.1	289	5	AA337010	AA337010
	24	258	10.1	469	155	R24760	yq43b06. r1
	25	257	10.0	369	169	BF818687	BF818687
	26	245	9.6	442	104	A1961078	A1961078
	27	238	9.3	912	173	BG107324	BG107324
	28	217	8.5	404	169	BF818347	BF818347
C	29	215	8.4	437	149	BF509823	BF509823
C	30	188	7.3	275	159	225140	HSB97F022. S
	31	163	6.4	518	3	AA210776	zr90d05. r
	32	144	5.6	152	143	BE062884	BE062884
	33	144	5.6	319	156	R89463	yq01b10. s1
	34	133	5.2	209	147	BF360596	BF360596
	35	129	5.0	408	147	BF351420	CM0-HT050
	36	120	4.7	131	113	AA297867	UI-H-BW0-
	37	118	4.6	396	31	AV652595	AV652595
	38	116	4.5	315	31	AV701438	AV701438
	39	113	4.4	543	31	AV703524	AV703524
	40	110	4.3	210	1	AA054853	F1-3207D
C	41	97	3.8	197	140	BE827340	CM2-ET001
	42	95	3.7	308	6	AA345403	EST51416
	43	91	3.6	851	145	BF209703	BF209703
	44	85	3.3	187	140	BE865366	BE865366
	45	85	3.3	188	140	BE865273	601014026

ALIGNMENTS

RESULT	1
AL514158	
LOCUS	
DEFINITION	AL514158 977 bp mRNA EST 13-FEB-2001
ACCESSION	AL514158 LIT_NF006.P12 Homo sapiens cDNA clone CLOBB005ZB08 5 prime mRNA sequence.
VERSION	AL514158
KEYWORDS	AL514158.1 GI:12777652
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 977) Li,W.B., Gruber,C., Jesse,J. and Polayes,D. Full-length cDNA libraries and normalization unpublished (2001)
AUTHORS	Contact: Genoscope
TITLE	Genoscope - Centre National de Sequencage
JOURNAL	Bp 191 91006 EVRY cedex - France
COMMENT	Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES	SOURCE
location/Qualifiers	
1. 977	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="Cl0B0052B08"	
/clone_1ib="LTI_NFL006_Pl2"	
/issue_type="Placenta"	
/note="Vector: pcMWSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pcMWSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"	
321 a 188 c 213 g 253 t 2 others	
BASE COUNT	
ORIGIN	

Query Match	35.0%	Score 896	DB 105	Length 977
Best Local Similarity	99.9%	Pred. No. 0		
Matches 946	Conservative	0	Mismatches 1	Indels 0
				Gaps
Oy 738	caactccctcattatcagaagaatcattgtaacccaacccgggaagctcgcagatgtaca	797		
Db 1	CACCTCCTCTGATTTACGAAAGTCATGTTTACCACACCGGAGGCTGGCAGATGTACA	60		
Oy 798	gttttatacaagtgtagcaacgcgttgagcatctgagcaagctatgtcatactgaagcaag	857		
Db 61	GTTTATTATCACAGGTGCAACCGCGGCGATTGGCAAGGATTATTCATTGAAACACAGAA	120		
Oy 858	gatgaggaacaaatattgtattatctgctgcgaagaacccggccagccacatccaaacttaagc	917		
Db 121	GATGGACCAAAATATTGTTATTTGGTCGAAGAACCGCCACCATTCACAAATTTCTAGCG	180		
Oy 918	acaatccatctgtctgtctgaagaatctgaagcagcttgaggaagaagcccttcgatact	977		
Db 181	ACAACTCTATCTGCTGCTGTGAAGAAATTGACAGCTTTGGAGGAAGGCCCTTGGCATTGTT	240		
Oy 978	gttgatgtgagagatgaacagcagatcagctgtctgcagctgtaggaagaagccatcaagaatt	1037		
Db 241	GTTGATGTGAGAGATGAAAGCAGATCAGTGGTGCAGTGGAGAAAGCCATCAAGAAATTT	300		
Oy 1038	ggaggaattgatattctgtgtaaatatgcagctgcacatagtttgaccaatacattgac	1097		
Db 301	GGAGGAATTTGATTTATTCGGTAAATATATGCGCCATGATTTTACCAATTCATTGGAC	360		
Oy 1098	acacctcccaagaagatttgatctgcagtatgaabaagtgaaacacagagagccactctgca	1157		
Db 361	ACACCTTCCAGAGATTTGGATCTGATGATGATGACGTGAACACAGAGGCACCTTACCTTCA	420		
Oy 1158	lctaaagcatglatctcccttcttgaagaagaagcaagcttgctcatatccctcaatacagt	1217		
Db 421	TCTTAAGATGTATTTCTTATTTGAAAGAGCAAGGTTGCTCATATCCTCAATTCAGT	480		
Oy 1218	ccaccactgaacctaaatccagcttggttccaacagcactgtgctataaccatctgctaa	1277		
Db 481	CCACACCTGAAACCTTAATTCAGTTTGGTTCCAAACACACCTGGCTTATACCTTCTAAG	540		
Oy 1278	tatgtatgtctatgtatgtctgtcttggaatgcagaagaatttaagaatgaatctgacgt	1337		
Db 541	TATGGTATGCTATGATGTATGTGCTTGGATATGCGAAGAAATTTAAAGGTGAATTCAGTC	600		
Oy 1338	aatgcattatgcccataaagaacccatacacactgtcctatgatatgctggaggaacct	1397		
Db 601	AATGCATTATGGCCTTAATAACAGCCATTAACACACTGCTGATGATATGCTGGAGGACCT	660		
Oy 1398	ggtatcgaagaacgcagctgtagaagaattgtatatactatgcagaatgcagatatcatcttc	1457		
Db 661	GGTATTCGAAACCCAGTGTACAAAAGTTGATATCATTTGACAGATGACGATATTCATTTTC	720		
Oy 1458	caaaagccaaaagaatttctactggaacctgtgcattgatagaataatcttaagaagaata	1517		
Db 721	CAAAAAGCCAAAAATTTTACTGTGGCACTTTGCTATTGATGAAGAAATATCTTAAACAGAA	780		
Oy 1518	ggaatagaagaatttcgacgcttatagtcaatataacacagctcatcccttgaaccagaattc	1577		
Db 781	GGATTAGCAAAATTTTGTACGTTTATGCAATTAACACAGCTCATCTTTTGCACAGATTC	840		
Oy 1578	tctctagatgaatacccaagaacagcttagcagaagaatggaatacactgtgtctgtcca	1637		
Db 841	TTTCTAGATGAATATCCAGAAAGCAGTTAGCAGAAAGATGATGAATCAACTGTCTGTCCA	900		
Oy 1638	gaatcagaagaagaagaactgcagctgcgaacccaacacagcttcggtg 1684			
Db 901	GAAATTCAAAGAGAAATGACAGCTGCAACCAACCAACCAAGCTTCGG 947			

RESULT 2

AL528484

LOCUS

DEFINITION

AL528484

861 bp

mrna

EST

13-FEB-2001

AL528484

U1_NF003_NBC3

Homo sapiens cDNA clone CSDDC029YE05

5

RESULT	2
LOCUS	AL528484
DEFINITION	AL528484 861 bp mRNA Homo sapiens cDNA clone CS00C029YE05 5

```

prime, mRNA sequence.
ACCESSION AL528484
VERSION AL528484.1 GI:12791977
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 861)
AUTHORS Li W.B., Gruber C., Jessee J. and Polayes D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr. Web : www.genoscope.cns.fr.

FEATURES
Source
1. 861
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSODC029YE05"
/clone_1b="LTI_NFL003_NBC3"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Peng Liang Life
Technologies, a division of Invitrogen 9600 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : filiang@life.com url :
http://fulllength.invitrogen.com"

BASE COUNT 270 a 172 c 188 g 226 t 5 others
ORIGIN
Query Match 31.3%; Score 802; DB 106; Length 861;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 802; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 atgttaaccaacacccggagcgtgacagatgtacagttttatcaacagtgcaagcgt 821
DB 60 ATGTTACCCACACCGCGGAGCTGGCAGATGTACAGTTTATCACAGTCCAGCCGT 119
QY 822 ggcatgggaagcattcattgtaagaagcgaagatgaggaagaattgtttgt 881
DB 120 GGCATTGGCAAGCTATTGCAATTGAAAGCAGCAAGATGAGCAATATGTTATTGCT 179
QY 882 gcaagaacccgacacacatccaactctcagcacatcatactgctgtaaga 941
DB 180 GCAAGACCGCCAGCAGCATCAAAACTTGAAGCAATCTATCTGCTGTAAGAA 239
QY 942 atggaagcgttgagggaagccttgccatgtatgttgatgagagatgaacagcag 1001
DB 240 ATTGAAGCGATTGGAGGAAGGCTTGCATGTATTGTTGATGTGAGATGAACGCGAG 299
QY 1002 atcaatgctgcagtgagagaagcattcaagaatttgaggaattgatatctgttaaat 1061
DB 300 ATCAGTGCCTGCGATGAGAGAAACCATCAAGAAATTTGGAGAAATGATATTTCTGTAAT 359
QY 1062 aatgcagtgatcattgattgacaaatatactgacacacccctacaagaagatgtagctg 1121
DB 360 AATGCGAGTGCATTTGTTGACATATACATGAGACACACATCAAGAGATTGGATGCG 419
QY 1122 atgatagaagtgaaacacccagagacactccttgatcctaaagatgtatccctatttg 1181
DB 420 ATGATGAAGCTAACCCAGAGGCACTTCTTCAATGATGTATTCCTTATTG 479
QY 1182 aaaaagagaagtgctcatatctcctaataatcagtcacacccctgaacctaataacagtt 1241

```

```

DB 480 AAAAGAGCAAAAGTTCCTATATCTCAATATCATGTCACCACTGAACCTAAATCCAGTT 539
QY 1242 tggttcaaaacagcactgigtctatataccattgctaagatgatatgatatgattgctt 1301
DB 540 TGGTTAAACAGCAGCTGCTTATACCATTTGCTAAATGATGATGCTTATGATGCTT 599
QY 1302 ggaatgcaagaagaatttaaaagtgaatctgacgtcaatgacgtctaaacacagc 1361
DB 600 GGAATGCGAAGAAATTTAAAGGTGAATTTGACAGTCAATGATTAATGGCTAAACAGCG 659
QY 1362 atacaacagctgctatgatatgctctggagagacccgtgatacgaagaagcgtgagaaa 1421
DB 660 ATACACACTGCTGTATGATATGCTGGAGAGCCGCTATCGAAAGCCAGTGTAGAAAA 719
QY 1422 gtgatatcattgcaagatgacgatatctcatttcacaaagccaaagatttctgac 1481
DB 720 GTTGATATATTCGATGACATGACGATATTCATTTTCCAAAGCCAAAGTTTACTGCG 779
QY 1482 aactgtcattgataaataatcttaaaagaagaagaagaatttgacgttatt 1541
DB 780 AACTTGTGATGATGAATAATCTTAAAGAAAGAAAGAAATTTGACGTTAT 839
QY 1542 gcaattaacacaggtcatcctt 1563
DB 840 GCAATTAACAGGTGATCCTT 861

RESULT 3
BE893726 602 bp mRNA EST 20-OCT-2000
LOCUS BE893726
DEFINITION 601366381 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3921920 5',
mRNA sequence.
ACCESSION BE893726
VERSION BE893726.1 GI:10355379
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 602)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1998)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
http://image.llnl.gov
Plate: LHM9755 row: k column: 09
High quality sequence stop: 602.

FEATURES
Source
1. 602
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3921920"
/clone_1b="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site:1. NotI;
Site:2. SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

BASE COUNT 196 a 109 c 129 g 168 t
ORIGIN
Query Match 22.8%; Score 584; DB 141; Length 602;
Best Local Similarity 100.0%; Pred. No. 8.6e-298;

```

Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1226 gaacctaataccaggttgggttcaaacagcctgtgttataccattgctagtgtat 1285
|||||
DB 19 GAACCTAAATCCAGTTGGTTCAAACAGCAGCTGTGCTTATACCATTCCTAAGTATGCTAT 78
1286 gctatgtatgtcttggaaatggcagaagaatttaaaagtgaatttcgaatgcat 1345
|||||
DB 79 GTTATGTATGTCTTGGAAATGGCAGAGAAATTTAAAGGTAATTCAGTCAATGCTATT 138
1346 atggcctaataacagcacaacacacgtcgtatgatgtatgtcgtggagacctgtatcga 1405
|||||
DB 139 ATGGCCCTAAACAGCCATACACAGCTGCTGTATGATGATATGCTGGAGAGACCTGATCGA 138
1406 aagcagctgtagaagaattgtatcatcattgcagatgcacatattccattccaagaac 1465
|||||
DB 199 AACCCAGTGTAGAAAGTTATATCATTCAGATGCGATGCGATATTCATTTCCAAAAAGCC 258
1466 aaaaagtttactgagcacttgcattgtatgaaataatcttaaaagaagaatagata 1525
|||||
DB 259 AAAAAGTTTACTGGCACTTTGTGATGATAAATATTTTAAAGACAGAGATATGA 318
1526 aaatttgacgttatagtcaatataaacaggtcacccttgcacacagattctctatga 1585
|||||
DB 319 AAATTTTGAAGCTTATGCAATTAACACAGGTATCTTTCACACAGATTTCTTTAGA 378
1586 tgaatacccaagaagcagttgagaagaagtgtgaatacaactgtgtgtcttcagaatcaa 1645
|||||
DB 379 TGAATCCCGAGAGCATTTGCAAGAAAGTGATCAACAGTGCTGTTCCAGAAATTC 438
1646 agagaagaacactgcagctgcaacaaacagcttctgagcgtgtggaagaagaattag 1705
|||||
DB 439 AGAAGAGAAACTCAGCTGCAACCAACCAACAGCTTCTGGAGCTGTGAAAGAACATTTAG 498
1706 aattgttaagagcctctcctcagatgatgtgtttaaagcaccactcaacatctatcgtt 1765
|||||
DB 499 AATTTGTTAAGAGCTCTCTCAGTGTATGTTGTTAAAGCCACATCAACATCATCTGTT 558
1766 tgaactctcagtggaagatgtgtgacagtggttcttgcattcga 1809
|||||
DB 559 TGAATCTCCGCTGGAAGATGTGGCAAGTGTGTTCTTGATCTGA 602

RESULT 4
LOCUS BF699109 887 bp mRNA EST 22-DEC-2000
DEFINITION 602126730F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4283671 5',
mRNA sequence.
ACCESSION BF699109
VERSION BF699109.1 GI:11984517
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 887)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CM1115 row: 1 column: 08
High quality sequence stop: 611.
Location/Qualifiers
1. 887

FEATURES
source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4283671"
/clone_11b="NIH_MGC_56"
/tissue_type="primitive neuroectoderm"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: brain; Vector: pMDR-LIB (Clontech); Site_1:
SfiI (ggcgagctggcc); Site_2: SfiI (ggcattatggcc);
Double-stranded cDNA was prepared from cell line RNA. 5'
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCCATTATGGC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGGCGGCGCGCCGACATG-dT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."

BASE COUNT 251 a 185 c 227 g 224 t

ORIGIN

Query Match 19.5%; Score 500; DB 168; Length 887;
Best Local Similarity 100.0%; Pred. No. 2.8e-253;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1114 tgatcgtatgatgaagtgtaaacaccagagcactccttgcattcattcaagcatgtattc 1173
|||||
DB 12 TGATCTGTATGATGAAAGTGAACCCAGAGGACCTTACCTTGATCTTAAAGATGATTTTC 71
1174 cttaattgaaagaagaagaattgtgctcatalcctcctaataatcagtcacacagccta 1233
|||||
DB 72 CTATTATTGAAAAAGCAAAAGTTGCTCATATCTCTCANTATCAGTCCAGCCTGAACCTAA 131
1234 atccagtttgcttcaaacagcactgtgctataccattgtgaagtatgtatgtatgt 1293
|||||
DB 132 ATCCAGTTTGCTTCAACACACAGCTGCTTATACCATGCTAAGTATGATGCTATGT 191
1294 atgtgcttgtaatgtgcagaagaatttaagtgtaattgcagtcacatgcatatgtgcta 1353
|||||
DB 192 ATGTGCTTGAATGGCGAAGAAATTTAAAGTGAATTCAGTCAACGATTTATGGCCTA 251
1354 aagcagcacaacacactgcgtgctatgatatagtctggaggaagcctgtgatacgaagcagt 1413
|||||
DB 252 AAACAGCCATACACACTGCTGCTATGATATGCTGGAGGACCTGGTATGAAAGCCAGT 311
1414 gtagaaagtgtatcatcattgcagatgcagatattccatttccaaagcacaagaatt 1473
|||||
DB 312 GTAGAAAGTTGATATCATTCGAGATGCACATATTCATTTCCAAAGCCAAAAAGTT 371
1474 ttaactggcaacttlytcatatgtatgaataatccttaagaagaagaataagaatttg 1533
|||||
DB 372 TTACTGGCAACTTGTTCATGATGAATAATCTTAAAGAAAGAAAGCAATAGAAAATTTTG 431
1534 acgttatatgaatttaaacccggtcatccttgcacaagaatttcttcttgaatgaatcc 1593
|||||
DB 432 ACGTTATGCAATTAACCGAGTCATCTTTGCAACCAAGATTTCTTATGATGAATACC 491
1594 cagaagcagttgacaagaata 1613
|||||
DB 492 CAGAGCAGTTTACCAAGAAA 511

RESULT 5
LOCUS AW152065/c 620 bp mRNA EST 03-NOV-1999
DEFINITION xf74a01.x1 NCL.GAP.Gas4 Homo sapiens cDNA clone IMAGE:2623752 3',
similar to WP:CI7610.8 CE16861 FAT-3: ALCOHOL DEHYDROGENASE ;, mRNA
sequence.
ACCESSION AW152065
VERSION AW152065.1 GI:6199963
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

OY	1554	ggatccctttgaaccagagttctctcttagatgaataccgaagcagttgtgcaagaa	1613
Db	301	GGTATCCCTTTGCAACCGATTTCTCTTAGATGAATACCCAGAAACAGTTTGGCAAGAAA	360
OY	1614	gtggaatccaactggtgctgttccagaatltcaagaagaagaaactgagctgcgaaccaaa	1673
Db	361	GTGGAATCAACACTGCTGCTGTTCCAGAAATTCAAAGAAAGAAACTGAGCTGCACCAAAA	420
OY	1674	ccacggttcctggaagctgtlgaagaagaacattagaattgtttaagactcttcagltatgat	1733
Db	421	CCACGTTCTTGAGAGCTGTGGAAAGAACATTAGAAATGTTAAGAGACTCTCTCAGTATGAT	480
OY	1734	gttgtttaa	1742
Db	481	GTGTGTTAAA	489

RESULT	7
LOCUS	AM966207
DEFINITION	AM966207 541 bp mRNA
ACCESSION	EST378280 MAGE resequences, MAGI Homo sapiens cDNA, mRNA sequence
VERSION	AM966207
KEYWORDS	EST.
SOURCE	EST. GI:8156043
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
	1 (bases 1 to 541)
	Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspar, R., Gay, C., Holm, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and Quackenbush, J.
TITLE	Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray
JOURNAL	Unpublished (2000)
COMMENT	Contact: John Quackenbush The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 3528 Fax: 301 838 0208 Email: johnd@tigr.org Plate: 224
FEATURES	Seq primer: Reverse.
source	Location/Qualifiers
	1..541
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone_lib="MAGE resequences, MAGI"
	/note="Vector: pBluescriptsm"
BASE COUNT	172 a 110 c 122 g 137 t
ORIGIN	

	Query Match	17.3%	Score 442	DB 122	Length 541
	Best Local Similarity	100.0%	Prod. NO. 1.5e-222		
	Matches 442	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	745	tcgatctagaagatcattgtaaccaacaccggaagctgycagatgtacagttlta	804		
Db	1	TCGTATCTACCAAAAGTCATGTATCCCAACACCGGAGCTGCAGATGTACAGTTT	60		
QY	805	tcaagatgacgaagccgtgcatlgtgcaaaagcattgcatltaaaacgaagaagtggag	864		
Db	61	TCACGGGGCAAGCCGTGTGCATGTGGCAAAAGCATTTGCAATGAACACAGCAAAAGATGGAG	120		
QY	865	caaatatgttatlgtlgtcaaaagaccgcacacatccaaactltagcacaaatc	924		
Db	121	CAAAATATTTGTTATTTGCTCCAAAGACCGCCGACCAATCCAAAAATTTCAGGCACAACT	180		
QY	925	atactgctgcgcgaagaatttgaagcaglttggaggaagggcttgcacatgatlgttatg	984		
Db	181	ATATGCTGCGCTGAAGAAATTTGAAGCATTTGGAGGAAGAGCCCTTGCCATCATATTGTTGATG	240		

QY	985	tggagagatgcaacgacgcatagtgtcgtcgcaatgtagaagacccatcaagaatttggagaa	1044
Db	241	TGAAGATGTGAACGACGAGATAGTGTGTCGAGTGGAGAAACCATCAAGAAATTTGGACGAA	300
QY	1045	ttagatattccggttaaataatgcccagtgccattgaatttgacatacatitggacacacta	1104
Db	301	TTGATATTTCGTGAATTAATAGCCAGTGCATTAAGTTTGAACCAATTCATTGGACACACTA	360
QY	1105	ccaagagattggaatcctgatatgaacgtagaacccagagacactactccttgatctaag	1164
Db	361	CCAAGAAATTGGATCTGCATGATGAAGGTGAACACGACGAGCCACTTTCGATCTTAAG	420
QY	1165	catgatcttcctatttggaaaa	1186
Db	421	CATGATATTCCTTAATTGAAAAA	442

RESULT	8
LOCUS	AA315514
DEFINITION	AA315514 487 bp mRNA EST187305 colon carcinoma (HCC) cell line II Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION	AA315514
VERSION	AA315514.1 GI:1967863
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. (Phases 1 to 487)
AUTHORS	Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult

TITLE	JOURNAL	MEDLINE	COMMENT
Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence	Nature 377 (6547 suppl), 3-174 (1995)	96026280	Other_ESTs: THC130362 Contact: Kerlavage, AR Bioinformatics The Institute for Genomic Research 9712 Medical Center Drive, Rockville, MD 20850 USA Tel: 3018699056 Fax: 3018699423 Email: arkerlavet@tigr.org For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (http://www.tigr.org/tdb/hgi/hgi.html) Seq primer: M13 Reverse.
Location/Qualifiers			
1.487			
/organism="Homo sapiens"			
/db_xref="ATCC (inhost):110493"			
/db_xref="taxon:9606"			
/clone_lib="Colon carcinoma (HCC) cell line II"			
/tissue_type="colon"			
/cell_type="KM12C"			
/cell_line="KM12C(HCC)-parental human colon carcinoma ; Dukes B2"			

Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
 Toshiyuki and Piero Carninci (RIKEN)
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
 Plate: LLM10036 row: f column: 14
 High quality sequence stop: 606.

FEATURES

source
 1. 829
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4373773"
 /clone_1ld="NIH_MGC_97"
 /lab_host="DH10B"
 /note="Organ: testis; Vector: pBluescript (modified
 pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTVN-3',
 size-selected for average insert size 2.2 kb and
 normalized to R0F 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."
 BASE COUNT 273 a 161 c 191 g 204 t
 ORIGIN

Query Match

Best Local Similarity 15.5%; Score 398; DB 171; Length 829;
 Matches 398; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1157 atccaaagcatgtatccctatcttgaaagaagaagtgctcatcctcaatcatcag 1216
 |||||
 Db 39 ATCTAAAGCATGTATTCCTTATTGAAAAAGAGCAAACTGCTCATATCCATATACG 98
 QY 1217 tccaccactgaacctaaatccagtttggttcaaacagcactgtgcttataccattgctaa 1276
 |||||
 Db 99 TCCACCACTGAACTGAATCCAGTTGGTTCAACAGCAGCTGCTTATACCATTTGCTAA 158
 QY 1277 gtatggtatgctatgatatgtgcttggaatgacagaagaatttaagtggaattgacgt 1336
 |||||
 Db 159 GTATGATGATGCTATGATGTGCTTGGAATGCGAAGAAATTTAAAGGTGAATTCAGT 218
 QY 1337 caatgcatatgagcctaaacacagcactatgctatgatatgctgagagagac 1396
 |||||
 Db 219 CAATGATTTATGGCTTAAACAGCCATACACAGCTGCTATGATATGCTGGGAGAGACC 278
 QY 1397 tggatcgaagaagcagtgtagaaaagttgatatcatcagatgcagatatccattt 1456
 |||||
 Db 279 TGGTATGGAAGCCAGTGTAAAGTTGATATCATTCGACATTCATTCATTTT 338
 QY 1457 ccaaaagccaaaagtttactgccaactgttcattgatataaattcttaagaaga 1516
 |||||
 Db 339 CCAAAAGCCAAAAGTTTACTGCAACTTGTGATTCATGATTAATAATCTTAAAGAGA 398
 QY 1517 aggaatagaanaatttgacgtttatgcaatlaaacag 1554
 |||||
 Db 399 AGGAATAGAAAATTTTGACGTTTATGCAATTAACACAG 436

RESULT 11
 AA305116 462 bp mRNA EST 18-APR-1997
 LOCUS EST176117 Colon carcinoma (Caco-2) cell line II Homo sapiens CDNA
 DEFINITION 5' end, mRNA sequence.
 ACCESSION AA305116
 VERSION AA305116.1 GI:1957444
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 462)

AUTHORS

Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
 C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
 O., Sutton,G., Blake,J.A., Brandon,R.C., Man'wal,C., Clayton,R.A.,
 Cline,T.R., Colton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
 L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodok,A.,
 Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,
 Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
 Moreno-Palanges,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
 Phillips,C.A., Ryder,S.E., Scott,J.L., Saudex,D.M., Shirley,R.,
 Small,K.V., Spriggs,T.A., Uterback,T.R., Weidman,J.F., Li,Y.,
 Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
 Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.,
 Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
 Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L.,
 Wei,Y.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Pannon
 M.R., Kosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
 Venter,J.C.

TITLE

Initial assessment of human gene diversity and expression patterns
 based upon 83 million nucleotides of cDNA sequence

JOURNAL

Nature 377 (6547 Suppl), 3-174 (1995)

MEDLINE

96026280

COMMENT

Other ESTs: THC130362

CONTACT

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

7912 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlavet@igrr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tdb/ngi/ngi.html>)

Seq primer: M13 Reverse.

FEATURES

Location/Qualifiers

SOURCE

1. 462

/organism="Homo sapiens"

/db_xref="ATCC (Inhost):128605"

/db_xref="taxon:9606"

/clone_1ld="Colon carcinoma (Caco-2) cell line II"

/sex="male"

/tissue_type="colon"

/cell_type="Caco-2"

/note="Organ: human colon adenocarcinoma; ATCC HTB 37"

/EcoRI: Site_2: XhoI"

BASE COUNT 145 a 96 c 110 g 111 t

ORIGIN

Query Match

Best Local Similarity 15.3%; Score 393; DB 5; Length 462;
 Matches 443; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 730 gtccgcgcacactcctctgatacgaagaatcatgttaccacaacgcggagctgacag 789
 |||||
 Db 1 GTCCGCCGCACCTCCTCGATCTAGCAAGATCAATGTTACCAACACCGGAGGCTGGCAG 60
 QY 790 gatgtacagttttatcacaggtgcaagcgtgtgcatgtggcaagcatatgcatgaag 849
 |||||
 Db 61 GATGTACAGTTTATATCACAGGTCAACCCCTGGCATTTGGCAAGCATTTTCATTGAAG 120
 QY 850 cagcaagaagtgaggaacaattgtttatgttgcaagaagcccgacgcacacatccaac 909
 |||||
 Db 121 CAGCAAGGATGAGCAAAATTTGTTATTCCTGCAAGACCGCCAGCCACATCCAAAC 180
 QY 910 ttctagacacatcatatctgctgtgaagaatggaagcagttggaaggaagccttgc 969
 |||||
 Db 181 TTCTAGGCAACAATCTATCTCTCTGTAAGAAATTTGAAGCAGTTGGAAGGAGGCTTGC 240

QY	970	catgattcgttgaatltgaaagatlgaaacacgaatcagtcgtcgaatgagggaaagccatca	1029
Db	241	CATGATTGTTGATGTGAAGATGAACACACAGTATGTCGTGCATGGAGAAAGGCATCA	300
QY	1030	agaaatcttgagagaaattgcatctctgtgtaataatgtccagttgcatlaqtlttgaccata	1089
Db	301	AGAAATTTGGAGGAATGTGATTTCTTGTAATAATATGCGAATGCGCAATTAGTTGACCAATA	360
QY	1090	catctgacacacctaccaaagagattgcatctgatatgtgaacgtgaaacaccagagccact	1149
Db	361	CATTGGACACACCTTACCAAGAGCTTGCTGTGATGTAAGCACTGGAACACACGAGGCACCT	420
QY	1150	accttcgcatctaaagcatgtatttc	1173
Db	421	ACCTTGCACTTAAGCATGTATTTC	444
RESULT	12		
LOCUS	W04752	484 bp	MRNA
DEFINITION	za42g10.r1 Soares fetal liver spleen 1mFLS Homo sapiens CDNA clone	EST	23-APR-1996
ACCESSION	W04752	IMAGE:295266 5'	W04752
VERSION	W04752.1	GI:1277679	EST
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
COMMENT	1 (bases 1 to 484) Hillier, L., Clark, N., Dubuque, T., Elliston, R., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rickin, L., Rolting, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Womdumnu, P. and Wilson, R. The WashU-Merck EST Project Unpublished (1995) Contact: Wilson RK Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.jnl.gov) for further information. Seq primer: mob. REGA+ET High quality sequence stop: 287. Location/Qualifiers 1..484		

BASE COUNT	166 a	67 c	77 g	172 t	2 others
ORIGIN					
Query Match	12.9%	Score 330	DB 157	Length 484	
Best Local Similarity	100.0%	Pred. No. 3.2e163			
Matches 330; Conservative	0	Mismatches 0	Indels 0	Gaps 0	

Oy	2056	aaaaaaaaagctgactgtgatctgagcccaaaaagtaaaaaagctcaacagcttataaaactaatg	2115
Db	1	AAAAAAAAAGTCGACTGCTTANGCTCAAAAAGTAAAAAAAGCTTAACAAGTTAAATCTAATG	60
Oy	2116	ctgtttctctctccgtctgatatataaagataagtcacgcttcttctggaaaaagataagat	2175
Db	61	TTTGTTTTCTTCTCTGTTATATTATTAAGAGTATGCACGTTTCTTGGAAGAAAGATAGAT	120
Oy	2176	tctgctctaaaaagacttgaaaatctgtaattaaaaatggcaagctaatcaacataaagcttca	2235
Db	121	TTGTCTCAAAAAGACTGGAATGTGAATTAAATGCAAGCTAATCAAAACATTAAGCTTCA	180
Oy	2236	ttaagtyggattctaaagacagctctgctgttttataattcaagggcttaacctctgagcc	2295
Db	181	TTAAGTGGGATTCTAACAACAGCTCGTGGTTTATATTTCMAAGGCTTTAAACCTTTGAGACC	240
Oy	2296	ttaacatctcaattcactgctcttctctccaagaaaaagtaatttggcgagaaagtcagtaag	2355
Db	241	TTTACATCTCATCTACCTCTCTCTCTCCCAAGAAAGATTTTGGCGGACAGTCAGATCAAG	300
Oy	2356	cagtaaaataagacgtcttcaaatctctctc	2385
Db	301	CAGTAATAATTAGCTCTTCCAAATCTTCTTGS	330

RESULT	13
LOCUS	R56767 482 bp mRNA EST 23-MAY-1995 R56767
DEFINITION	y95fa10.t1 Soares Infant brain INIB Homo sapiens cDNA clone IMAGE:41102.5', mRNA sequence.
ACCESSION	R56767
VERSION	R56767.1 GI:826873
KEYWORDS	EST.
SOURCE	human. Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 482) Hallier,L., Clark,N., Dubnque,T., Elliston,K., Hawkins,M., Holman ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rickin,I., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston ,R., Williamson,A., Woldmann,P. and Wilson,R. The Washu-Merck EST Project Unpublished (1995)
TITLE	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu Insert Size: 3243 High quality sequence stops: 326 Source: IMAGE Consortium, LNL This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 3243 Std Error: 0.00 Seq primer: M13RP1 High quality sequence stop: 326.
JOURNAL	
COMMENT	Location/Qualifiers 1..482 /organism="Homo sapiens" /db_xref="GDB:413643" /db_xref="taxon:9606" /clone="IMAGE:41102" /clone_lib="Soares Infant brain INIB" /sex="female" /dev_stage="73 days post natal" /lab_host="DH10B (ampicillin resistant)" /note="Organ: whole brain; Vector: ltafmd BA; Site_1: Not I ; Site_2: Hind III; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACGGAGAAATTCGCCGCCAGCAATTTTTTTTTTTT 3']: double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned

into the Not I and Hind III sites of the lafmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo.
BASE COUNT 150 a 64 c 89 g 172 t 7 others
ORIGIN

Query Match 12.7%; Score 326; DB 156; Length 482;
Best Local Similarity 100.0%; Pred. No. 4,2e-161;
Matches 326; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2137 ttcaagagatagcagcttctgctgaaagagaattgtctctcaagaacttgaa 2196
|||||
Db 1 TTATAAGGATATGCACCTTGTCTGGAAGAAGATAGATTGCTCTTAAAGACTTGAA 60
QY 2197 ttgtaattaaaatggaagcattcaacaataagcttcatttaagtggattcctaagacag 2256
|||||
Db 61 TTGTAATTAATAATGCAAGCAAGTCAATCAAAACATPACCTTCATTAGTGGGATCTAAGACAG 120
QY 2257 tctgtgtttataatcctaaggttttaaccttgagccttaactcattcactgtgctt 2316
|||||
Db 121 TCTGTGTTTATATATTCAGAGGTTTAACTTGTAGCCTTACATCTCATTCACCTGCTT 180
QY 2317 tctccaagaaagatatttggcgagacagtcagatcaagcagtaaaattiaactcttcaa 2376
|||||
Db 181 TCTCCAGAAAGATATTTGGGGCGACAGTCAGATCAAGCAAGTAAATTACTCTTTCAA 240
QY 2377 atctctctgtcattgaaagcagtcagctgtctttaaattttagtttgattgtat 2436
|||||
Db 241 ATCTTCTTGTATATTAATGCAAGTCAAGTCTGTTTAAATTTTGTAGTTGGATTGTAT 300
QY 2437 actaatgaataacttaactgattctt 2462
|||||
Db 301 ACTAATGAAATCTTAATGATGTTT 326

RESULT 14
AA622988/c 386 bp mRNA EST 21-OCT-1997
LOCUS np58H04.s1 NCI_CGAP_Br2 Homo sapiens cDNA clone IMAGE:1130551 3'
DEFINITION similar to WP:CI7G10.8 CE02490 ALCOHOL DEHYDROGENASE ;, mRNA
sequence.

ACCESSION AA622988
VERSION AA622988
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 386)
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
Natlional Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmett-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/bdrip/image/image.html
Insert Length: 937 Std Error: 0.00
Seq primer: -40ml3 fwd. EM from Amersham
High quality sequence stop: 342.
Location/Qualifiers

FEATURES
source 1. 386
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1130551"

/clone.lib="NCI_CGAP_Br2"
/sex="female, pooled"
/tissue_type="breast"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from pooled bulk
breast tumor tissue, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT7T3
vector. This library is the normalized version of
NCI CGAP Br1.1. Library was constructed by Bento Soares
and M. Fatima Bonaldo "

BASE COUNT 99 a 92 c 58 g 137 t
ORIGIN

Query Match 12.7%; Score 325; DB 9; Length 386;
Best Local Similarity 100.0%; Pred. No. 1,4e-160;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1694 agaacattgaattgttgaagactctcagtgatgattgtttaaagccactcaagc 1753
|||||
Db 386 AGAAACATTTGAATGTTAAGACTCTCAGTGATGATGTTTAAAGCCACTCAAGC 327
QY 1754 aatctatctgttgaactctcagtgaaagatgtagcagtggttcttgatctgaaaag 1813
|||||
Db 326 AATCTATCTGTTTGAACCTCCGGTGAAGATGTGGCCGTGTGTTCTTGATCTGAAAG 267
QY 1814 caaggtggaatgctcgatataagagaccttcgatacgaagcagatgtagatgagat 1873
|||||
Db 266 CAAGGTGGAATGTCGATATGAGAGCCCTTCGATGAGCAATGCGATGAGATAT 207
QY 1874 gactactgtagactctgttgaagatttccaggaacaaacaaacatgagcattcat 1933
|||||
Db 206 GACTACTGATGACTTGTGAAATAATGTTTCAGGGAACCAACCAATGCATTCAT 147
QY 1934 gtccaggaattgaagattgaactatgacatgcccctgcaatgaattggagagcattat 1993
|||||
Db 146 GTCCAGGGAATTTGAAGATTAAAGTAAACATGCCCTTACCAATTAATTTGGAGAGCTAAT 87
QY 1994 gaatcagatgaatgccagactgtga 2018
|||||
Db 86 GAATCAGATGAATGCCAGACTGTGA 62

RESULT 15
Z19446 340 bp mRNA EST 10-FEB-1993
LOCUS HSB28P112 STRATGENE Human skeletal muscle cDNA library, cat.
DEFINITION #936215. Homo sapiens cDNA clone 28P11, mRNA sequence.
ACCESSION Z19446
VERSION Z19446.1 GI:29287
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 340)
AUTHORS M.D., Duprat,S., Hougatte,R., Juneau,M.N., Lamy,B., Lorenzo,F.,
Sedastiani,Kabakchis,C. and Tessier,A.

TITLE IMAGE: molecular integration of the analysis of the human genome
and its expression
C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
JOURNAL 95277534
MEDLINE
COMMENT Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698

Email: genexpress@genethon.fr
strand(+), single read
Warning!! There has been some controversy in the scientific
literature over
possible contamination of this library by yeast and prokaryotic
sequences.
See Science 259:1677-1678 (1993).

FEATURES
source
1. 340
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="GDB:D05690E"
/db_xref="taxon:9606"
/clone="28f11"
/clone_lib="STRATAGENE Human skeletal muscle cDNA library,
cat. #936215."
/sex="female"
/tissue_type="skeletal muscle"
/dev_stage="19 years"
/note="Organ: leg muscle; Vector: Lambda ZAPII; Tissue
from female, 19 years old, normal leg muscle. Cloning
vector is Lambda ZAPII, in vivo excision from lambda ZAPII
to phuscript SK(+). Genexpress library reference is B. "
BASE COUNT 106 a 47 c 59 g 127 t 1 others
ORIGIN

Query Match 12.4%; Score 317; DB 159; Length 340;
Best Local Similarity 100.0%; Pred. No. 2.5e-156;

Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2128 cctgtatattttaaggaatgacagcttctgtcttggaaaagatagaatttgcctctaaa.2187
|||||
Db 1 CCTGTTATTTATTAAGGATATGCACGTTGTTCTGGAAAAGATAGATTGTCCTTAAAA 60
OY 2188 gacttgaatttgaattaaatggaagcaatcaacaataagcttcattaaagtggatt.2247
|||||
Db 61 GACTTGAATTGTAATTAATGCGACAGTATCAACATAGCTTCATTAGTGGGATT 120
OY 2248 ctaagaagctcgtgttttataattcaagggttaaccccttgagccttacaatcatt.2307
|||||
Db 121 CTAAAGACACTGCTGTTTATTTATTTCAAGGGTTTAACTTTGACCTTACATCTCATTT 180
OY 2308 cactgtcttccccaagaaagtattttggggagacagtcagatcaagcagtaaaattag.2367
|||||
Db 181 CACTGCTTTCTCCAGAAAGATTTTGGCGGACAGTCAGATCAAGCAGTAAAAATTAG 240
OY 2368 ctcttcaaatctcttgcacatglaaaatgaagctagctgttttaaaattttaatt.2427
|||||
Db 241 CTCTTCAATCTCTGTCATGTAATAATGAAGCTAGTCTGTTTAAATTTTACTTTT 300
OY 2428 ggactgtataactaataga.2444
|||||
Db 301 GGATTGTATATCAATGA 317

Search completed: June 12, 2001, 13:10:48
Job time: 5425 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 12, 2001, 10:40:42 : Search time 169.81 Seconds
(without alignments)
8804.334 Million cell updates/sec

Title: US-09-464-039-8

Perfect score: 2561
Sequence: 1 aggcagagatgcaaacga.....gsgmggraswmwawrwammc 2561

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapect 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

N_Geneseq_0401.*
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	556.2	21.7	4951	19	V27206 Vector p1rAP3 DNA
2	400	15.6	633	18	T97159 SR alpha enhancer/
3	354.2	13.8	633	19	V32786 Sequence of the sp
4	259.4	10.1	755	10	N92604 Recombinant DNA ve
5	259.4	10.1	9045	17	T42902 DNA sequence which
6	259.4	10.1	9047	5	N40080 Sequence complemen
7	249.4	9.7	720	10	N90825 Human T lymphotrop
8	249.4	9.7	720	15	O68398 HTLV-1 LTR genomic
9	249.4	9.7	720	18	T47852 HTLV-1 long termin
10	230	9.0	1675	14	O50836 HTLV-1 p21x cDNA.
11	230	9.0	1866	14	O50835 HTLV-1 tax/rex spl

ALIGNMENTS

HindIII fragment 0
Human secreted pro
Plasmid pMRG3169 e
MMP9 promoter GFP
Vector plasmid pCM
AAV vector sequenc
AAV vector sequenc
AAV vector sequenc
Plasmid pMRG3196 e
Nucleotide sequenc
MMP9 promoter beta
DNA sequence of p1
Shuttle vector pad
PSAB132. Syntheti
Shuttle vector pad
Recombinant adenov
Plasmid pAV.CMVlac
AV.CMVlac2 cis pla
Vector containing
Expression vector
Expression vector
Ad.AV.CMVlac2 hybr
CDNA sequence of p
CDNA sequence of p
CDNA sequence of p
Plasmid pADCMVgag-
Shuttle vector pad
Recombinant trans-
EST clone AR253.
MO 9923223 Seq ID
MO 9923223 Seq ID
Plasmid pSV17.ID.L
Expression vector

12 181.4 7.1 422 13 025258
13 180.4 7.0 198 21 C08959
14 167.6 6.5 7287 19 V02042
15 166.2 6.5 5068 20 X84027
16 166 6.5 3796 21 A27831
17 166 6.5 6253 20 X08454
18 166 6.5 6280 20 X08455
19 166 6.5 6280 20 X08456
20 166 6.5 6295 19 V02043
21 166 6.5 6981 21 X45931
22 166 6.5 7380 20 X84028
23 166 6.5 7612 21 Z39629
24 166 6.5 7852 17 T27556
25 166 6.5 7892 13 Q30906
26 166 6.5 7897 17 T27555
27 166 6.5 8236 18 T47202
28 166 6.5 8509 18 T59271
29 166 6.5 8509 20 X33862
30 166 6.5 9077 19 V09005
31 166 6.5 9077 19 V09006
32 166 6.5 9077 21 Z45251
33 166 6.5 9077 21 Z45253
34 166 6.5 10398 17 T15286
35 166 6.5 10398 19 V22130
36 166 6.5 10398 19 V22139
37 166 6.5 10398 19 V22139
38 166 6.5 12135 18 T60555
39 166 6.5 19307 17 T27558
40 166 6.5 36538 18 V86402
41 164.6 6.4 304 20 V86402
42 163.6 6.4 7286 20 X57224
43 163.6 6.4 7938 20 X57235
44 163 6.4 5141 16 T51451
45 163 6.4 5141 16 Q94114

RESULT 1
ID V27206 standard; cDNA; 4951 BP.
XX V27206;
AC 12-OCT-1998 (first entry)
DT Vector p1rAP3 DNA sequence.
DE
XX
XX
KW Secreted protein; membrane-associated protein; protein secretion;
KW signal peptide; alkaline phosphatase; p1rAP3; vector; reporter; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FT mat_peptide 1313..2782
FT /*tag= a /product= human placental alkaline phosphatase
FT
FT
XX
XX MO9822491-A1.
XX
XX 28-MAY-1998.
XX
XX 06-NOV-1997; 97WO-US20201.
XX
XX 19-NOV-1996; 96US-0752307.
XX
XX (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX
XX Gearing DP, Levinson DA, McCarthy SA;
XX WPI: 1998-312407/27.
XX
XX P-PSDB: W55047.

xx Identifying DNA encoding mammalian protein having signal sequence.
 pt by cloning in reporter gene vector without signal sequence, -
 pt transforming bacteria and mammalian cells, and identifying cells
 pn that express reporter
 xx
 xx Claim 6; Fig. 2; 50pp; English.
 ps

This is the DNA sequence of vector p1rAp3. This mammalian expression vector contains a cDNA encoding human placental alkaline phosphatase (AP, see W55047) lacking a signal sequence. When p1rAp3 is transfected into a mammalian cell line, such as COS7 cells, AP protein is neither expressed nor secreted since the AP cDNA of p1rAp3 does not encode a signal peptide or a membrane anchor sequence. However, insertion of a cDNA encoding a signal peptide sequence into p1rAp3 facilitates the expression and secretion of AP upon transfection of the DNA into mammalian cells. The presence of AP activity in the supernatants of transfected cells therefore indicates the presence of a signal sequence in the cDNA of interest. This forms the basis of a novel method for the identification of genes encoding novel proteins having a signal sequence, i.e. secreted or membrane-associated proteins of potential therapeutic value. An isolated clone (see V27707), designated eth0018f2, was isolated using the novel method and encodes a human neural adhesion protein (see W55045) protein having multiple, consecutive IgG domains. The method is very sensitive and is suitable for high throughput screening techniques and automation.

SQ Sequence 4951 BP; 1151 A; 1410 C; 1303 G; 1087 T; 0 other;

Query Match	21.7%	Score	556.2	DB	19	Length	4951
Best Local Similarity	96.2%	Pred. NO.	1.8e-78				
Matches	656	Conservative	0	Mismatches	14	Indels	12
						Gaps	8

OY	1	aggcgaagatctgcgaagcatgcgctctccaattctgtagcgaacacatagctccggccct	60
Db	131	aggcgaagatctgcgaagcatgcgctctccaattctgtagcgaacacatagctccggccct	187
OY	61	aactccgcgcacatcccgcccccactccgncnccagctccgcgcacatctctccgccatgagc	120
Db	188	aactccgcgcacatcccgcccccactccgncnccagctccgcgcacatctctccgccatgagc	245
OY	121	tgaactaatctttctatcttaattcagcaga-gccgagacgcgcttcgcgcctctgagctatccag	175
Db	246	tgaactaatctttctatcttaattcagcaga-gccgagacgcgcttcgcgcctctgagctatccag	300
OY	180	aagtagctgaggaaggctttttgagagcgctagcgttttgcaaaaagccctc-cgatcgaag	233
Db	306	aagtagctgaggaaggctttttgagagcgctagcgttttgcaaaaagccctcctccgcatacgaag	365
OY	239	ggtccgcatactctctctcttcacacgcgcgcgcgcgcctacactgagagccgcacatccacgcgct	298
Db	366	ggtccgcatactctctctctcttcacacgcgcgcgcgcctacactgagagccgcacatccacgcgct	422
OY	299	gagtcgcgtcttcgcgcgcctccgcgcctgctgagctcctcctgaaactgctgcgcgcgtctagat	356
Db	426	gagtcgcgtcttcgcgcgcctccgcgcctgctgagctcctcctgaaactgctgcgcgcgtctagat	485
OY	359	aagtttaaaagttcaagtgctgagaaacggggcgtcttgcgcgcgcgtctcccttgaggagcctaactag	418
Db	486	aagtttaaaagttcaagtgctgagaaacggggcgtcttgcgcgcgcgtctcccttgaggagcctaactag	545
OY	419	aactcagccgcgtctcccaagccttgcgcgcgcgtctgctcacaactcagctcttgcttc	478
Db	546	aactcagccgcgtctcccaagccttgcgcgcgcgtctgctcacaactcagctcttgcttc	605
OY	479	agtttctcgtctcgcgcgcgtctacagatccaaagctctgaaaaacagaagaagttaactgta	538
Db	606	-gcttctcgtctcgcgcgcgtctacagatccaaagctctgaaaaacagaagaagttaactgta	666
OY	539	agtttagtcttcttgctcttataattcaaggtcccgagatccgg---tggtggtgtaaatca	594

Db 665 agtttagctcttttgccttcaattccaggctcccgatccggtgatccaaatct 724

Oy 555 aagaactgctcccaagtgatgtgtgccttcactctagagccgtacgaagtgcttactc 654

Db 725 aagaactgctcccaagtgatgtgtgccttcaactctagagccgtacgaagtgcttactc 784

Oy 655 tgccttaagaagctgcggaattc 676

Db 785 tgccttaagaagctgcggaattc 806

RESULT	2
T97159	
ID	T97159 standard; DNA; 633 BP.

AC T97159;

DT 11-MAY-1998 (first entry)
VY

DE SR alpha enhancer/promoter.
XY

KW Vaccine; B-cell lymphoma; leukaemia; tumour;
gene amplification; immunohistochemistry; SV40; ex-

XX KW enhancer; ds.

OS Chimeric = Human T cell leukaemia virus type 1
US Chimeric = Khesus macaque polyoma virus.

XX
PN W09741244-A1

06-NOV-1997
XX
PD

25-APR-1997

AA
PR 06-DEC-1996: 96HIS-0761277

PK 01-MAY-1996; 96US-0644664.
XX
XX

XX PA (GENT-) GENTLOPE CORP:

P1 Deirney DM
XX

DR WFL; 1991-349143/50
XX

PT comprises at least 2 different recombinant variable regions of

XX
F1
immunoglobulin molecules derived from B cell lymphoma cells

Example 1; page 104; 1/1pp; English
PS
XX

CC This DNA sequence comprises the HindIII/XhoI fragment of the
CC SR alpha enhancer/promoter in plasmid pCCL-SR alpha 296. The
CC SR alpha enhancer/promoter is composed of human T cell leukemia
CC virus 1.5' untranslated sequences and the SV40 enhancer. It is
CC reported to increase expression from the SV40 enhancer/promoter by
CC 10-fold in most cells, and is active in a broad range of cell
CC types. The SR alpha enhancer/promoter has been utilised in
CC expression vectors designed for efficient expression of genes in
CC eukaryotic cells. The invention provides vectors and improved
CC methods for the expression and co-amplification of genes encoding
CC recombinant proteins in cultured cells. The methods permit the
CC isolation of cell lines which have co-amplified input recombinant
CC sequences which encode an amplifiable marker, one or more
CC expression vectors encoding a protein of interest and optionally a
CC selectable marker. The amplified cells provide large quantities
CC of recombinant proteins suitable for immunotherapy for treatment of
CC lymphomas and leukemias. The methods permit the production of
CC custom vaccines, including multivalent vaccines that reflect the
CC degree of somatic variation found in a patient's tumour.

SQ Sequence 633 BP; 114 A; 210 C; 155 G; 154 T; 0 other;

	Query Match	15.6%	Score 400:	DB 18:	Length 633;	
	Best Local Similarity	96.1%:	Pred. No. 3.2e-54;			
	Matches 487:	Conservative 0:	Mismatches 6;	Indels 14;	Gaps 7:	
Oy	1 aggcagaagatgtcgaacgatcatctcaaatattgtagcgaaccataagtccggccct 60					
Dd	131 aggcagaagatgtcgaacgatcatctc-aattgtagcgc-aacctatgcc-gccct 187					
Oy	61 aactcgcccatcccgcgccttaactcgnccagtttcgggccattctccggccatggc 120					
Dd	188 aactcgcccatcccgcgccttaactcgcgc--agttccggccattctccggccatggc 245					
Oy	121 tgactaatcttcttatttatgacaga-gccgagccgcgcctggcctctgagctattccag 179					
Dd	246 tgcactaatcttcttatttatgacagaaggccgagccgcctcgccctctgagctattccag 305					
Oy	180 aagtgtagagagagcttttttgaaagcctaggccttttgaaaaagctcctgatatcgagg 239					
Dd	306 aagtgtagagagagcctttcttgaaagcctaggccttttgaaaaagcctcctcgag----- 358					
Oy	240 gctcgacatctcctctcaacgcgcgcgcgcgcgccttaactgagagccgacatcacgcggttg 299					
Dd	359 gctcgacatctcctctcaacgcgcgcgcgcgcgccttaactgagagccgacatcacgcggttg 418					
Oy	300 agtcgagctctcgcgcctccgcctctgtgtgctcctctgaactcgtlccgcgcltagyta 359					
Dd	419 agtcgagctctcgcgcctccgcctctgtgtgctcctctgaactcgtlccgcgcltagyta 478					
Oy	360 agttaaagctcaggtcgaagacgggagcctttctccggcgctccctctgtaggactactaga 419					
Dd	479 agttaaagctcaggtcgaagacgggagcctttctccggcgctccctctgtaggactactaga 538					
Oy	420 ctccagccggcctccacagccttgcctgaaccctgcttgcctcaactcagctcttgtttcca 479					
Dd	539 ctccagccggcctccacagccttgcctgaaccctgcttgcctcaactcagctcttgtttcc- 597					
Oy	480 gtcttcgtctctcgcgcgttlacagalc 506					
Dd	598 gtcttcgtctctcgcgcgttlacagalc 624					
	RESULT 3					
ID	V35786					
XX	V35786 standard; DNA; 633 BP.					
AC	V35786;					
DT	22-SEP-1998 (first entry)					
DE	Sequence of the specification.					
KW	Induce; apoptosis; guanine phosphoribosyltransferase; gpt; luciferase;					
KM	cyclokinine; tumour necrosis factor; interleukin-1; inhibitory effect;					
OS	Intracellular signal transmission; ss.					
XX	Unidentified.					
PN	WO9822578-A1.					
PD	28-MAY-1998.					
PF	12-NOV-1997; 97MO-JP04126.					
PR	15-NOV-1996; 96JP-0305163.					
PA	(CYTO-) INST CYTOSIGNAL RES INC.					
PI	Nagasawa Y, Yoshida H;					
DR	WPI: 1998-312464/27.					
PT	Test system for detecting intra-cellular signal transmission					
TT	Inhibition - using vector containing apoptosis-inhibiting or					

[illegible]

Feature	Location/Qualifiers
Key	1..353
LTR	/*tag= a /label= U5 354..381
LTR	/*tag= b /label= R 582..755
LTR	/*tag= c /label= U3
JP01277489-A.	
07-NOV-1989.	
28-APR-1988;	88JP-0105677.
28-APR-1988;	88JP-0105677.
(GANK) GAN KENRYUKAI ZH.	
WPI: 1989-368596/50.	
Recombinant DNA vector - contains HTLV-I long terminal repeat ligated to part of U5, for increased chloramphenicol acetyltransferase activity.	
Disclousure; ; Japanese.	
HTLV-I LTR is inserted into pSV-CAT, the U3 region is lost, and the R-U5 region gives 10-100 fold CAT expression. Vector can be integrated in a wide range of animal cells for high levels of expression.	
Sequence 755 BP; 161 A; 255 C; 178 G; 161 T; 0 other;	
Query Match	10.1%; Score 259.4; DB 10; Length 755;
Best Local Similarity	99.3%; Pred. No. 2.7e-32;
Matches 271; Conservative 0; Mismatches 1; Indels 1; Gaps 1	
237	ggagctcgacatctctccttcacagcgccgcgcctactgagcgccatccacagccgg 296
352	ggggctcgacatctctccttcacagcgccgcgcctactgagcgccatccacagccgg 411
297	ctgagctcgatctctcgcgcctccgcgcctggtggtgctctcctgaactgagcgccgttag 356
412	ctgagctcgatctctcgcgcctccgcgcctggtggtgctctcctgaactgagcgccgttag 471
357	gtaagtttaaaagtcaggtcagagacgggaccttgcgcgcgtccctctgagagctaacct 416
472	gtaagtttaaaagtcaggtcagagacgggaccttgcgcgcgtccctctgagagctaacct 531
417	agactcagcgccgcttcacagcttgcgtgacctgacctgacctgaactgaacttcttgg 476
532	agactcagcgccgcttcacagcttgcgtgacctgacctgacctgaactgaacttcttgg 591
477	tcagtttctcgtcttcgcgcgtttacagatccaa 509
592	tc-gttctcgtcttcgcgcgtttacagatccaa 623
RESULT 5	
T42902	
T42902 standard; DNA; 9045 BP.	
T42902;	
16-JUN-1997 (first entry)	
DNA sequence which regulates expression in HTLV and HIV.	
gene expression; regulation; plasmid; viral infection;	
human T-cell leukemia; HIV; antiviral agent; detection; cancer;	
gene therapy; USP; ss.	

XX	Homo sapiens.
OS	
XX	
FH	Key
FT	misc_feature
FT	Location/Qualifiers
FT	1..757
FT	/tag= A
FT	/label= LTR
FT	polyA_signal
FT	8584..8589
FT	/*tag= b
FT	misc_feature
FT	8278..9032
FT	/*tag= c
XX	
PN	M09630522-A1.
XX	
PD	03-OCT-1996.
XX	
PF	19-MAR-1996;
XX	96WO-JP00719.
XX	
PR	27-APR-1995;
PR	95JP-0104299.
PR	24-MAR-1995;
XX	95JP-0066559.
PA	(SHIO) SHIONOGI & CO LTD.
XX	
PI	Igarashi H, Okumura K, Orita S, Saiga A, Sakaguchi G;
DR	
XX	WPI: 1996-455367/45.
PT	DNA molecule with gene expression regulation activity - for use in
PT	e.g. treatment of human T-cell leukaemia and HIV, as antiviral agent
PT	and for detecting cancer
PS	Claim 2; Page 38-43; 77pp; Japanese.
XX	
CC	This sequence represents a DNA molecule with gene expression regulation
CC	activity. This sequence is used in a plasmid for regulation of gene
CC	expression, and treatment of viral infection pref. human T-cell leukemia
CC	and HIV. The plasmid also encodes a protein which is used as an antiviral
CC	agent, and also in a method for detecting cancer. The DNA molecule and
CC	protein have potential uses in gene therapy, and the plasmid may also
CC	have potential use in the treatment of TSP.
XX	
SQ	Sequence 9045 BP; 2086 A; 3162 C; 1713 G; 2084 T; 0 other;
<hr/>	
Query Match	10.1%; Score 259.4; DB: 17; Length 9045;
Best Local Similarity	99.38; Pred. NO. 3.5e-32;
Matches 271; Conservative	0; Mismatches 1; Indels 1; Gaps 1;
OY	237 ggggctgcacatcctcccttcacagcgccgccgtactaggccgcacacacgcgg 296
DB	
OY	352 ggggctgcacatcctcccttcacagcgccgccgtactaggccgcacacacgcgg 411
OY	297 ttgagtcgcttctgcgcgtccgcgccttggtggtcctctgaactgogtccgcgcttag 356
DB	
OY	412 ttgagtcgcttctgcgcgtccgcgccttggtggtcctctgaactgogtccgcgcttag 471
OY	357 gtaagttaaagaagccaagttcgaacggcgacctgttcgcggcgcccaccttgaaagctaact 416
DB	
OY	472 gtaagttaaagaagccaagttcgaacggcgacctgttcgcggcgcccaccttgaaagctaact 531
OY	417 agaatcacgcgcgctctccacagctttgacctgaaccctgtctgtcaactatactatttgtt 476
DB	
OY	532 agaatcacgcgcgctctccacagctttgacctgaaccctgtctgtcaactatactatttgtt 591
OY	477 tcaagttctgttcttgcgcggttacagatcaa 509
DB	
OY	592 tc-gtttctgttcttgcgcggttacagatcaa 623
DB	
<hr/>	
RESULT	6
N40080	
ID	N40080 standard; cDNA: 9047 BP.
XX	

PR 18-DEC-1991; 91JP-0354839.
XX
PA (SHIO) SHIONOGI & CO LTD.
XX
DR WPI; 1993-338952/43.
XX
PT Oligo-nucleotide primer - for detecting mRNA of human T cell
XX leukemia virus 1 by polymerase chain reaction
XX
PS Disclosure; Fig 7; 19pp; Japanese.
XX
XX (Q50835) shows cDNA prepared from RNA extracted from a HTLV-1
CC infected cell. This was amplified by PCR and was used to construct
CC p21x cDNA (Q50836). Primers and probes were then manufactured
CC (Q50837-43). these are useful for the detection of HTLV-1 infection
XX
SQ Sequence 1866 BP; 394 A; 678 C; 386 G; 408 T; 0 other;

	Query March Best Local Similarity Matches	230: Conservative	9.0% 100.0% 0;	Score 230: Pred. No. 1.1e-27; Mismatches	DB 14: 0;	Length 1866; Indels	0;	Gaps
QY	237	ggggtcgcacatctctctctcaacgagccgcgcgcgcctacatgagcgcgcgcacacgcgcg	296					
Db	1637	ggggctcgcatctctctctctcaacgagccgcgcgcgcctacatgagcgcgcgcacacgcgcg	1696					
QY	297	ttgagtcgcgtctctcgccgcctccgcgcctctgtgagcctcttgaaatcggtctccgccttaag	356					
Db	1697	ctgagtcgcgtctctcgccgcctccgcgcctctgtgagcctcttgaaatcggtctccgccttaag	1756					
QY	357	gtaagctttaaagctcaagctcagagtcagacggcgagcctctgtctcgcgagctcccttgagacctact	416					
Db	1757	gtaagctttaaagctcaagctcagagtcagacggcgagcctctgtctcgcgagctcccttgagacctact	1816					
QY	417	agactacgcgcgcgtctccaagctttgacgcctctgtgcacactctca	466					
Db	1817	agactacgcgcgcgtctccaagctttgacgcctctgtgcacactctca	1866					

```

RESULT 12
025258
ID Q25258 standard; DNA; 422 bp.
XX
AC Q25258;
XX
DT 26-NOV-1992 (first entry)
XX
DE HindIII fragment of pSEL.
XX
KW Cytokine; plasmid pSEL; HTLV-1; human T-lymphocyte virus;
KW mouse; alpha-globin; E.coli cloning vector; ds.
XX
OS Synthetic.
XX
PN EP488900-A.
XX
PD 03-JUN-1992.
XX
PF 29-NOV-1991; 91EP-0403243.
XX
PR 29-NOV-1990; 90FR-0014961.
XX
PA (ERAP ) ELF SANOFT.
PA (SNET ) SANOFT SA.
XX
PI Caput D, Ferrara P, Miloux B, Minty A, Vita N;
XX
DR WPI; 1992-185765/23.
XX
PT New monocyte chemoattractive cytokine - for treatment of cancer
PT and parasitic infections, e.g. leishmaniasis, leprosy or Chagas
PT disease

```

XX This fragment contains a sequence close to the 5'-UTR of
PS HTLV-1 and the mouse alpha-globin distal intron. The fragment is
XX used in the construction of plasmid pSE1 which was designed as a
CC cloning vector in E.coli and an expression vector in animal cells.
CC See also Q25259-Q25262.
XX
SQ Sequence 422 BP; 66 A; 142 C; 121 G; 93 T; 0 other;

	Query Match	7.1%	Score 181.4	DB 13	Length 422
	Best Local Similarity	85.6%	Pred. No. 3.7e-20		
	Matches	237	Conservative	0	Mismatches 36; Indels 4; Gaps 3
QY	239	ggtctgcatctctctctcaagcgccgcgcgcctcaactatgagcgccatccacgcgcggt	298		
Db	5	ggctcgcatctctctctctcaagcgccgcgcgcctcaactatgagcgccatccacgcgcg	63		
QY	299	gagtcggtcttcgcgcgcctcccgccctgtgtgtgcctctcaactgaactgtctgcgcgt	358		
Db	64	gagtcggtcttcgcgcgcctcccgccctgtgtgtgcctctcaactgaactgtctgcgcgt	123		
QY	359	aagtttaagctcagtcgagacgggccttgtccgcgcgcctcccttgagcctactctg	418		
Db	124	aggtccccaag--gtagcgccgagacaagggccgggtctctgaccttgagccttaactactctg	181		
QY	419	actaagcgcgatctctcaacgcttctgtcgtgacgtgtctgtctcaacttaagctcttgcttc	478		
Db	182	actaagcgcgatctctcaacgcttctgtcgtgacgtgtctgtctcaacttaagctcttgcttc	241		
QY	479	agttctctgtctctgcgcgcttacaatccaagctctg	515		
Db	242	gtttctctgtctctgcgcgcttacaactccaagctctg	277		

RESULT	13
C08959	
ID	C08959 standard; cDNA; 198 bp.
XX	
AC	C08959;
XX	
DT	06-OCT-2000 (first entry)
XX	
DE	Human secreted protein 5' EST, SEQ ID NO: 13034.
XX	
KW	Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation
KW	gene therapy; chromosome mapping; ss.
XX	
OS	Homo sapiens.
XX	
PN	EP1033401-A2.
XX	
PD	06-SEP-2000.
XX	
PF	21-FEB-2000; 2000EP-0200610.
XX	
PR	26-FEB-1999; 99US-0122487.
XX	
PA	(GEST) GENSET.
XX	
PI	Dumas Milne Edwards J, Duclert A, Giordano J;
XX	
DR	WPI: 2000-500381/45.
XX	
PT	New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT	obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT	diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX	
PS	Claim 1; SEQ ID 13034; 71pp + CD-ROM; English.
XX	
CC	The present sequence is one of a large number of 5' ESTs derived from


```

ID X84027 standard; DNA; 5068 BP.
XX
AC X84027;
XX
DT 26-AUG-1999 (first entry)
XX
DE MMP9 promoter GFP construct.
XX
KM Reporter gene; skin metabolism promoter; matrix metalloproteinase 9;
KM MMP9 promoter; green fluorescent protein; skin treatment; ss.
XX
OS Synthetic.
XX
PN WO9303743-A1.
XX
PD 24-JUN-1999.
XX
PF 30-APR-1998; 98WO-US08794.
XX
PR 17-DEC-1997; 97US-0069945.
XX
PA (GEHO ) GEN HOSPITAL CORP.
PA (SHIS ) SHISEIDO CO LTD.
XX
PI Amano S, Burgess RE, Ehama R, Kishimoto J, Nishiyama T;
XX
DR MPI; 1999-395095/33.
XX
PT Transgenic animal having a reporter gene coupled to a skin
PT metabolism-related promoter
XX
PS Disclosure; Fig 3; 62pp; English.
XX
CC This sequence represents a construct containing the matrix
CC metalloproteinase 9 (MMP9) promoter linked to the green fluorescent
CC protein (GFP) coding sequence.
CC The invention relates to a non-human transgenic animal having a reporter
CC gene coupled to a skin metabolism promoter. The transgenic animal can be
CC used in a method for evaluating a treatment for its effect on skin. The
CC method, together with the transgenic animal are useful for evaluating a
CC compound for its effect on the health or appearance of the skin. The
CC method enables rapid and efficient evaluation of compounds for their
CC effect on skin.
XX
SQ Sequence 5068 BP; 1205 A; 1393 C; 1362 G; 1108 T; 0 other;

Query Match 6.5%; Score 166.2; DB 20; Length 5068;
Best Local Similarity 83.3%; Pred. No. 1.1e-17;
Matches 189; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 505 tccaagctctgaaacagaagaagtaactgtaagtttaagctcttctcttatttc 564
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 813 tcgaggaactaaacacagaagaagtaactgtaagtttaagctcttctcttatttc 872

QY 565 aggtcccgatccggtgtgtgtgcaaatcaagaactgctcctcagtgatgtgcctt 624
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 873 aggtcccgatccggtgtgtgtgcaaatcaagaactgctcctcagtgatgtgcctt 932

QY 625 actctagcgctgtaaggaagtgttactctgtccttaaaagctcggaattctaatacga 684
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 933 actctagcgctgtaaggaagtgttactctgtccttaaaagctcggaattctaatacga 992

QY 685 ctcaactaagggtgtgacccacgctgcgcgcgcgcgcgcgtgt 731
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 993 gccgcaattcccggtcgccaccacatggtgagcaaggcgaggaagctgt 1039

```

Search completed: June 12, 2001, 11:52:00
 Job time: 4278 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 12, 2001, 09:24:42 ; Search time 1765.19 Seconds
(Without alignments)
12674.640 Million cell updates/sec

Title: US-09-464-039-8

Perfect score: 2561

Sequence: 1 aggcagaagatgcagaagca.....gsgmgtasmmwawrrammc 2561

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est48:*
 118: gb_est49:*
 119: gb_est50:*
 120: gb_est51:*
 121: gb_est52:*
 122: gb_est53:*
 123: gb_est54:*
 124: gb_est55:*
 125: gb_est56:*
 126: gb_est57:*
 127: gb_est58:*
 128: gb_est59:*
 129: gb_est60:*
 130: gb_est61:*
 131: gb_est62:*
 132: gb_est63:*
 133: gb_est64:*
 134: gb_est65:*
 135: gb_est66:*
 136: gb_est67:*
 137: gb_est68:*
 138: gb_est69:*
 139: gb_est70:*
 140: gb_est71:*
 141: gb_est72:*
 142: gb_est73:*
 143: gb_est74:*
 144: gb_est75:*
 145: gb_est76:*
 146: gb_est77:*
 147: gb_est78:*
 148: gb_est79:*
 149: gb_est80:*
 150: gb_est81:*
 151: gb_est82:*
 152: gb_est83:*
 153: gb_est84:*
 154: gb_est85:*
 155: gb_est86:*
 156: gb_est87:*
 157: gb_est88:*
 158: gb_est89:*
 159: gb_est90:*
 160: gb_est91:*
 161: gb_est92:*
 162: gb_est93:*
 163: gb_est94:*
 164: gb_est95:*
 165: gb_est96:*
 166: gb_est97:*
 167: gb_est98:*
 168: gb_est99:*
 169: gb_est100:*
 170: gb_est101:*
 171: gb_est102:*
 172: gb_est103:*
 173: gb_est104:*
 174: gb_est105:*
 175: gb_est106:*
 176: gb_est107:*
 177: gb_est108:*
 178: gb_est109:*
 179: gb_est110:*
 180: gb_est111:*
 181: gb_est112:*
 182: gb_est113:*
 183: gb_est114:*
 184: gb_est115:*
 185: gb_est116:*
 186: gb_est117:*
 187: gb_est118:*
 188: gb_est119:*
 189: gb_est120:*

190: em_gss_pin1:*
 191: em_gss_pin2:*
 192: em_gss_pro:*
 193: em_gss_rtd1:*
 194: em_gss_rtd2:*
 195: em_gss_rtd3:*
 196: em_gss_rtd4:*
 197: em_gss_rtd5:*
 198: em_gss_vrt1:*
 199: em_gss_vrt2:*
 200: em_gss_vrt3:*
 201: em_gss_vrt4:*
 202: em_gss_vrt5:*
 203: em_gss_vrt6:*
 204: em_gss_vrt7:*
 205: em_gss_vrt8:*
 206: em_gss_vrt9:*
 207: em_gss_vrt10:*
 208: em_gss_vrt11:*
 209: em_gss_vrt12:*
 210: em_gss_vrt13:*
 211: em_gss_vrt14:*
 212: em_gss_vrt15:*
 213: em_gss_vrt16:*
 214: em_gss_vrt17:*
 215: em_gss_vrt18:*
 216: em_gss_vrt19:*
 217: em_gss_vrt20:*
 218: em_gss_vrt21:*
 219: em_gss_vrt22:*
 220: em_gss_vrt23:*
 221: em_gss_vrt24:*
 222: em_gss_vrt25:*
 223: em_gss_vrt26:*
 224: em_gss_vrt27:*
 225: em_gss_vrt28:*
 226: em_gss_vrt29:*
 227: em_gss_vrt30:*
 228: em_gss_vrt31:*
 229: em_gss_vrt32:*
 230: em_gss_vrt33:*
 231: em_gss_vrt34:*
 232: em_gss_vrt35:*
 233: em_gss_vrt36:*
 234: em_gss_vrt37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	963.8	37.6	977	105	AL514158
2	835.6	32.6	861	106	AL528484
3	636.8	24.9	692	113	AM299683
4	604.4	23.6	963	141	BE892079
5	596.6	23.3	767	18	AI316248
6	590	23.0	602	141	BE893726
7	581	22.7	620	111	AM152065
8	571.4	22.3	887	168	BE699109
9	523.6	20.4	541	122	AM966207
10	518.8	20.3	675	144	BE151259
11	515.6	20.1	815	145	BE182478
12	512.2	20.0	693	110	AM044841
13	506.2	19.8	889	141	BE914941
14	492.8	19.2	693	165	BE307752
15	485.4	19.0	618	174	BE245008
16	473.2	18.5	938	144	BE101858
17	470.4	18.4	487	5	AA315514
18	461	18.0	640	21	AI526513

19	449.4	17.5	462	5	AA305116
20	448.2	17.5	497	32	AV750182
21	446.8	17.4	616	32	AV714646
22	429	16.8	526	117	AW619441
23	409.4	16.0	544	8	AA475560
24	408.2	15.9	510	102	AT787419
25	404.2	15.8	624	144	BF138546
26	398.4	15.6	829	171	BF979524
27	386.8	15.1	437	149	BF509823
28	386.2	15.1	731	165	BE309622
29	384.6	15.0	482	156	R56767
30	384.4	15.0	442	104	AT961078
31	383	15.0	538	8	AA475551
32	378.6	14.8	530	104	AT956634
33	374.4	14.6	484	157	W04752
34	368.4	14.4	527	32	AW012259
35	365.6	14.3	468	165	BE305849
36	356.4	13.9	466	3	AA213310
37	353.6	13.8	441	163	BE119891
38	350.6	13.7	386	9	AA622988
39	347.2	13.6	508	13	AA881807
40	346.8	13.5	474	14	AA986652
41	340.2	13.3	426	169	BF818728
42	339.6	13.3	483	113	AM259182
43	338.2	13.2	447	156	R89462
44	336.8	13.2	887	145	BE238534
45	334.2	13.0	625	166	BE372615

ALIGNMENTS

RESULT 1
 LOCUS AL514158 977 bp mRNA EST 13-FEB-2001
 DEFINITION AL514158 LFI_NFL006_PL2 Homo sapiens cDNA clone CLOB0052B08 5
 ACCESSION AL514158
 VERSION AL514158.1 GI:12777652
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 977)
 AUTHORS Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
 source
 1. 977
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CLOB0052B08"
 /clone_id="LFI_NFL006_PL2"
 /tissue_type="placenta"
 /note="Vector: pCMVSPORT 6; site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com"

BASE COUNT 321 a 108 c 213 g 253 t 2 others

Query Match 37.6% Score 963.8; DB 105; Length 977;
 Best Local Similarity 99.7%; Pred. NO. 3,3e-156;
 Matches 974; Conservative 2; Mismatches 0; Indels 1; Gaps 1;

Qy	738	cacctccctgcatctacgaagtcgtgttaccaccccggaagcctgagatgtaca	797
Db	1	CACCTCCCTGATCTACGAAGTCATGTTACCAACCGGAGGCTGCGAGATGTACA	60
Qy	798	gtttatcacaggtgcgaagcctgtgcattgcgaagctatgtcatgtgaagcgaag	857
Db	61	GTTTTATCACAGGTGCGAAGCCTGTGCGATTGCAAGCTATTGCAATGAAGCAGAAAG	120
Qy	858	gatgaggaataatgtttatgtctgcgaagaccccgcccaactccaactcttcgac	917
Db	121	GATGAGGAATAATGTTTATGTGTCGAAGACCCGCCACACATCCAAACCTTCAGGC	180
Qy	918	acaatctactgtctgcgaagaattggaacagttggaagaagccttcacatgtatt	977
Db	181	ACAATCTACTGCTGCTGCGAAGAAATTCAGCACTTGGAGAAAGCCTTCGATGATT	240
Qy	978	gttgatgtgagagatgcgaacagatcagtgctgcagctggaagaagccatcaagaatt	1037
Db	241	GTTGATGTGAGAGATGGAACAGCATGATGCTGCTGAGAGAGCATCATCAAGAAATT	300
Qy	1038	ggaggaattgtatctcgtgtaataatgacagctgcattagtttgaccaataattggac	1097
Db	301	GGAGGAATTGATTCGTGCTGAATAATGCGCATGCTCCATTATTTGACCAATATTTGAC	360
Qy	1098	acacctcaagagatgtgatctgatgtgaagctggaacacaggaagcaactacttgca	1157
Db	361	ACACCTTACCAAGATTTGATGATCTGATGATGACCTGAACACCAAGGACCTTACTTGCA	420
Qy	1158	tctaaagcatgtatctccttatttgaaaaagagcaaaatgtgcatatcctcctaatacagt	1217
Db	421	TCTAAAGCATGTATTCCTTATTGAAAAAGCAAGCTGCTCATTCCTCAATATCAGT	480
Qy	1218	ccaccactgaaccttaaatccagttgtgttcaaaagacagctgtgttataccattgctaag	1277
Db	481	CCACCACCTGAACCTTAATCCAGTTTGTTCNAACGCACTGTGCTTATTCACCATTCGTAAG	540
Qy	1278	tatgtatgtctatgtatgtcttggaatgagcagaagaatttgaagctgaaattgcagtc	1337
Db	541	TATGTATGTCTATGTATGTATGTCTTGGAATGCAAGATTTAAGTGAAATTCGATC	600
Qy	1338	aatgcattatgaccttaaaacagccatacacacgtcgtctatgtatgtcgtggagacct	1397
Db	601	AATGCATTATGCTCTTAATAAACGACCATACACCTGCTATGATGATGCTGGAGGACCT	660
Qy	1398	ggtatgaaagcagtggttgaagaagttgatatcatgtgagatgagatatctcatcttc	1457
Db	661	GGTATGCAAAAGCCAGTGAGAAAGTTGATATTCATTCGATGAGCATATTTCCATTTTC	720
Qy	1458	caaaagccaaaagcttctacgtgcaactgtgcatatgtgaataatcttaaaagaa	1517
Db	721	CAAAAGCCAAAAGCTTTTACGCACTTGTCTATGATGAATATCTTAAAGAAAGAA	780
Qy	1518	ggaatagaaatcttgacggtttatgtcaatlaaaccaaggtcaatctctggaacagatttc	1577
Db	781	GGAATAGAAATTTTGAAGTTTATGCAATATTAACAGAGTCTTCCTTGGCAACGATTTTC	840
Qy	1578	ttcttagatgaatccagaagcagtttagcaagaagaattgataactatggtgcgttcca	1637
Db	841	TTCTTAGTGAATCCCAAGAGCAGTTAGCAAGAGAGGAGATCACTGGGCTGTTCCA	900
Qy	1638	gaattcaaaagagaagaaactgcaagctgcaaaccaaaacagcttcgg-agctgtggaaga	1696
Db	901	GAAATTCAAAGAGAAAGTGCACACTGCACCAACCAAAACACAGCTTGTGATGAGAGA	960
Qy	1697	aacattagaattgtta 1713	
Db	961	DACATTTAGAAATGTGTA 977	

RESULT	2	LOCUS	AL528484	861 bp	mRNA	EST	13-FEB-2001
DEFINITION	AL528484	LT1.NFL003_NBC3	Homo sapiens	CDNA clone	CSDDC029YE05	5	
ACCESSION	AL528484						
VERSION	AL528484.1	GI:12791977					
KEYWORDS	EST.						
SOURCE	human.						
ORGANISM	Homo sapiens						
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.						
AUTHORS	Li, W.B., Gruber, C., Jessee, J., and Polayes, D.						
TITLE	Full-length CDNA libraries and normalization						
JOURNAL	Unpublished (2001)						
COMMENT	Contact: Genoscope						
FEATURES	Genoscope - Centre National de Sequencage						
source	BP 191 91006 EVRI cedex - France						
	Email: sequence@genoscope.cns.fr , Web : www.genoscope.cns.fr .						
	Location/Qualifiers						
	1..861						
	/organism="Homo sapiens"						
	/db_xref="taxon:9606"						
	/clone="CSDDC029YE05"						
	/clone_1lb="LT1.NFL003_NBC3"						
	/sex="male"						
	/tissue_type="neuroblastoma cells"						
	/lab_host="DH10B"						
	/note="Organ: brain; Vector: PCMVSPORT 6; 1st strand CDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded CDNA was digested with NotI and cloned into the NotI and EcoRV sites of the PCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : li.fang@life.com URL : http://fulllength.invitrogen.com "						
BASE COUNT	270 a	172 c	188 g	226 t	5 others		
ORIGIN							
Query Match	32.6%	Score 835.6	DB 106	Length 861			
Best Local Similarity	99.4%	Pred. NO. 3.8e-134					
Matches 847	Conservative 2	Mismatches 1	Indels 2	Gaps 1			
QY	714	gctgcgcgcgcgcgtctgtccgcacacctcctctgatactag--aaagtcattaccca	771				
Db	10	gctgcgcgcgcgcgcgtctgtccgcacacctcctctgatactag--aaagtcattaccca	69				
QY	772	acacccggagagctctgacagagatgatacagtttatacacaaggctgacagcgttgcatctgca	831				
Db	70	acacccggagagctctgacagagatgatacagtttatacacaaggctgacagcgttgcatctgca	129				
QY	832	aagctatgcatctgaaagcagaagaagatgtagaacaatactgttactgtctgtaagaagacagc	891				
Db	130	aagctatgcatctgaaagcagaagaagatgtagaacaatactgttactgtctgtaagaagacagc	189				
QY	892	cccagccacatccaaacattcctagggcacaacatctatactgctgctgtaagaagacagc	951				
Db	190	cccagccacatccaaacattcctagggcacaacatctatactgctgctgtaagaagacagc	249				
QY	952	ttggaggaagaagcctctgcatatctgtctgtgtagagagatgaacaagcagatcagctgctg	1011				
Db	250	ttggaggaagaagcctctgcatatctgtctgtgtagagagatgaacaagcagatcagctgctg	309				
QY	1012	cagtgagaaagccatcacaagaacattctgagagatcgatatctctgttaataatgcagctg	1071				
Db	310	cagtgagaaagccatcacaagaacattctgagagatcgatatctctgttaataatgcagctg	369				
QY	1072	ccattagtttgaccaatatacatctggaacacacccaagaagatctgatactgatactgaagc	1131				

Db	370	CCATTGATGTTGACCAATTCATGTGGACACAGCCTACCACAAGAATTTGGATCTCATGATGAACG	429
Qy	1132	tgaacaccagaggcaacctactctgtcatcctaagaatgtaattcccttatttgaaaaagagca	1191
Db	430	TGAACACACAGAGAGCACCTTGCATGATTAAGCATGTATTCCTTATTTGAAAAAGAGCA	489
Qy	1192	aagtgcctcatacccataatatcagtcacaacactgaaacctaatccagtttgtgtccaac	1251
Db	490	AGTGTCCTCATTCCTCAATATCACTGCCACACCTGAACCTTAATCGACTTTGGTTCAAC	549
Qy	1252	agcaactgtgcttatcacatttgttaagtatgltatgtcattgtatgtctgtgaatgtag	1311
Db	550	AGCACTGTGCTTATTCACATTCGTAAAGTATGTAATCTCATGTATGCTTGAATGGCAG	609
Qy	1312	aagaatttaaagtgtaaaattgcagtcacatgtcatatggccctaaaacagccatacacatg	1371
Db	610	AMGATTTTAAAGCTGAATTCGACAGTCATTCGATTTATGGCCCTAAAACACCCATACACATG	669
Qy	1372	ctgctatcgatatactgtggaggagacctgtgatccgaagccagtgtagaaagtgtatatca	1431
Db	670	CCTGCTATGATGATTCGTGGAGAGCACTCGTATTCGAACCCAGCTGTAGAAMAATTTATATCA	729
Qy	1432	ttagcagatgcagacatattccatcttccgaagaagccaaaagctttacitggcgaactttca	1491
Db	730	TTGCGATGTCACCATATTCATTTTCCAAAAGCCAAAAGTTTACTGCAACTTTGTCA	789
Qy	1492	ttagatgaaataatctctaaagaagaaggaataagaataatttgacgtttatgcaatltaaac	1551
Db	790	TTGATGAATAAATTCCTTAAAGAAGAAGAAATGAAATTTTAGCGTTATTCGAATTAAC	849
Qy	1552	caggtcalccctt 1563	
Db	850	CAGGTCAATCCTT 861	
RESULT 3			
LOCUS	AW299683/C		
DEFINITION	AW299683 692 bp mRNA EST 18-JAN-2000 x542e02.x1 NCI-CGAP.Kid1l Homo sapiens cDNA clone IMAGE:2722314 3'		
ACCESSION	AW299683		
VERSION	AW299683		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 692) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)		
AUTHORS	Contact: Robert Strausberg, Ph.D. Tel.: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution Information can be found through the I.M.A.G.E. Consortium/LNL at: www.bio.lnl.gov/bdip/image/Image.html Seq primer: -40Up from GIDco High quality sequence stop: 466.		
JOURNAL	location/Qualifiers		
COMMENT	1..692		
FEATURES	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2722314" /clone_id="NCI-CGAP.Kid1l" /lab_host="DH10B"		
source			

OY	1734	gtttttt-aaagcaatcaaacatcatctcg-cttgaactccggtaagaatgtgtgca	1791
Db	481	gttttttAAAGCCACATCAACCATCTATCTGTTTAACTCCGGTAAATGGTGGCA	540
OY	1792	cgtgtttctttgatcttgaagaaggtgtggaaatgtcggatatyagaagcccttgatc	1851
Db	541	CGTGGTTTCTGTGATCTGAAACAAAGGAGGCGCATCTCGAATATGAGAGCGCTTCTGAT	600
OY	1852	agg---cagaatgtgtgatcatagtga-tyactactgaatgaattgtataaaatgtttccagg	1907
Db	601	CAGGCCAGATGTGTGTATGATAGATTGACTACTAGATGACTTTGTTAACTGCTTTTC--GG	658
OY	1908	aaactaaaccaacaatbgtcatctatctatgtcgaaggaatltgaagatt--aaagtataatcgt	1965
Db	659	GAACTTAAACCAACAATATGCATCTCTGTCAAGGACCCTCGAGATACACAGTAACATATGG	718
OY	1966	ccctagacaatcaaatgtggagaagctataga	1995
Db	719	GCTTAGCAGCTAGTGGGAGAAAGTATGATA	748

LOCUS	767 bp	MRNA	EST	17-DEC-1998
DEFINITION	uJ26G04.y1	Sugano mouse kidney mK1a Mus musculus cDNA clone		
IMAGE	1921110	5' similar to WP:C45B11.3	CE05427	ALCOHOL
DEHYDROGENASE	;	mrna sequence.		

```

INCESSION  AT040240
VERSION     AT1316248.1  GT:4031515
REVISION    1

```

SOURCE
ORGANISM

REFERENCE
AUTHORS
1 (bases 1 to 767)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.

TITLE	The Washu-HHMT Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	Contact: Marra M/Mouse EST Project

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:977402

seq primer: custom primer used
High quality sequence stop: 483.

FEATURES
source

```

1. .767
   /organism="Mus_musculus"
   /strain="C57BL"
   /db_xref="taxon:10090"
   /clone="IMAGE:1921110"
   /clone_1fp="Sugano mouse kidney mkla"
   /sex="female"
   /dev_stage="adult"
   /lab_host="DH10B"
   /note="Organ: kidney; Vector: pME18s-FL3; Site_1: DraIII
(CACGCTGTG); Site_2: DraII (CACCACTGTG); 1st strand cDNA
was primed with an oligo(dT) primer
AATGGCCATTCTTTTCTTTTTTTTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [ATTGGCGCTACTGCI] digested
and cloned into distinct DraIII sites of the pME18s-FL3
vector (5' site CACGCTGTG, 3' site CACCACTGTG). XhoI should
be used to isolate the cDNA insert. Size selection was

```

performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTTCGCTCTAAAGCTCG and 3' end primer CGACCTCGAGCTCAGCACA."

Query Match	23.3%	Score 596.6	DB 18	Length 767
Best Local Similarity	87.0%	Pred. No. 4.8e+93		
Matches 667	Conservative 0	Mismatches 99	Indels 1	Gaps 1

Dy 760 tcaagttacccaacaccgggagygctgtgcagaatytacagttttcttcacaagtgtaaaccc 819
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 TCATGCTGCCAACACACTGGAGGCTAGCGAGATGCACAGTTTTCACACAGTGCAAGCC 60

Qy 820 gggcattggcaagcattcattggaagcagcaagagctggagcaaatctgtcattg 879

Db 61 GAGGCATTGGCGAAGCTATTGCTTGAAGCCGATAGATGGAGCCAAATTGTGCTATTG 1200

Oy 880 ctgcgaagaccgccacgcaatccaaaactcttagcgacaatcctaactgtgctgaag 939
||||| ||||| ||| || ||||| ||| ||||| ||| ||||| |||
Db 121 CTGCGAAGACCAACCCTGCGCACAATCTCTGGCACATCTACACCGATGCTGAAG 180

QY 940 aaatctgaagcagcttggagaagaagcccttcacatgatatctgtgatgtagagataatgaacagc 999
|||||
Db 181 AAATTTGAACAGCTGGAGGAGCAGCGCCCTTGTCTCGTTGATGTGAGAGATGAACAGC 240
|||||

QY 1000 agacacgtctgcagtcgagagaagccatcaagaattctgaggaattgatcatctcgtaa 105

Db 241 ATATCAACACTGCAGTCGAGAGAACTGTGGAGAAATTTTGGAGAAATTGATTTTTGGTGA 300

Qy 1060 ataatgcagtcgcatcattagtcttgaccacatatcattggacacacctaaccaaggatgtgac 111
+ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 301 ACAATGCCAGTGTCTATTAGCTTGACCAACACAGCTGGGACTCCGAGCCAAGAAGAGTAGTACT 360

Qy 1120 tgatgatgaacgtgaacacccagagcaaccttgcatacctaagaatgtatttccttatt 117
Db 361 TGTATGATGATGTGTGACACCGAGGGGACCTACCTTACATCCAAAGCATGTATTTCTTTT 420

DY 1180 tgaanaagacaaagtgtcctatatccctaatactagtcaccactgaacctaatatccag 1233

DB 421 TAAATAAAGACCAACATGCTGTATTATTTTCCAAATTCATCGCCGCCCATCCTTAATACCCG 490

Oy 1240 ttctggttcaaacagcactgtgcttataccattgcctaagtatgtagtctatgtatgtgc 129
Ob | |||||
A81 tccaccccccacacacacaccccccccccccccccccccccccccccccccccccccc E40

QY 1300 ttggaatgcgagaagaattttaaagtcgaatctgcagtcacatgcattatggcctaacaag 135

QY 1360 ccatacacactgcctcgtatgatatgctggtgagagaccctggtatcgaagaagccagtctgaag 141

Qy 1420 aagttgatacatctgcagatgcagacatatccatttccaaaaagccaaaaagtttactg 147
||||||| ||||||| || ||||||||||||||| || || |||||||||||||

1480 gcaactctgcattgatgaata-cttaaaagaagaagataga 1525

RESULT 6

LOCUS	BE893726	602 bp	mRNA	EST	20-OCT-2000
DEFINITION	601436638F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3921920 5'				
mRNA sequence.					

ACCESSION	BE033720
VERSION	BE893726.1
	GI:10355379

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 602)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC/DCPD/DRP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
plate: L1AM9755 row: k column: 09
High quality sequence stop: 602.
Location/Qualifiers
1. 602
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3921920"
/clone_lib="NIH-MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

BASE COUNT 196 a 109 c 129 g 168 t
ORIGIN

Query Match 23.0%; Score 590; DB 141; Length 602;
Best Local Similarity 99.8%; Pred. No. 6.6e-92;
Matches 601; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1209 aatacagtcacacact-gaacctaaatccagttggttcaaacagcactgtgtatcac 1267
|||||
DB 1 AATATCACTCCACCCGTCGAAATCAATCCAGTTGGTTCAACAGCACTGGCTTATAC 60
|||||
QY 1268 catgtcagaatgtagtctatgtagtctgttggaatgagcagaagaattaaagtga 1327
|||||
DB 61 CATTGCTAGATGATGTATGTATGTCTGCTTGGAATGCGAAGAATTTAAAGTGA 120
|||||
QY 1328 aattgcacatgcatatgagcctaaacagccatcacactgtgtatgatatgt 1387
|||||
DB 121 AATTGCACTCAATGATATGCTTAACACGCCATACACACTGCTGTATGATATGCT 180
|||||
QY 1388 gggagacgtgtatcgaagccagtgtagaagaatgatcatctgtagatgcagaca 1447
|||||
DB 181 GGGAGACCTGGTATCGAAGCCAGTGTAGAAAGTTGATATCGAAGATCGACATA 240
|||||
QY 1448 ttcaatttccaaaagccaaaagtttactgagcaacttgcattgtagaataatctt 1507
|||||
DB 241 TTCCATTTTCCAAAAGCCAAAAGCTTTACGCAACTTTGTCATGTGATAAATATCTT 300
|||||
QY 1508 aaaagaagaagaatagaatatttgagcttattgcaatlaaacaggtcatcctttga 1567
|||||
DB 301 AAAAGACAGAGAAATGAAAATTTTGACGTTTATGCAATTTAAACCGGTCTATCCCTTGA 360
|||||
QY 1568 accagattcttctagaatgatacccgaaagcagttgagaagaagtggaaatcaactgg 1627
|||||
DB 361 ACCGATTTCTTTAGATGAATACCGAAGCACTTGAAGAAGATGGATATCACTGG 420
|||||
QY 1628 tgcgtgtccagaatcagaagaagaacatgcagctgcaaccaaaacacagtlctggagc 1687
|||||
DB 421 TGCTGTCCAGAAATTCAGAAAGAGAACTGCACCTGCACCAAAAACACGCTTGGAGC 480
|||||

QY 1688 tgtggaagaacatttaaatgttaagactctcagtgatgtagtgtgtaagccac 1747
|||||
DB 481 TGTGGAAGAACATTTAAGATTTGAGACCTCTCAGTATGATGTTGTTAAAGCCAC 540
|||||
QY 1748 tcaagcaatctatcgtttgaactccgcgtgaagatgtagtggatcttctgacatc 1807
|||||
DB 541 TCAACCATCTATCTGTTGAACTCTCCGCTGAACATGCTGACAGCTGTTCTTGATCT 600
|||||
QY 1808 ga 1809
||
DB 601 GA 602
|||||

RESULT 7
AM152065/c 620 bp mRNA EST 03-NOV-1999
LOCUS x174a01.x1 NCI-CGAP-Gas4 Homo sapiens CDNA clone IMAGE:2623752 3'
DEFINITION similar to WP:G17G10.8 CEI16861 FAT-3: ALCOHOL DEHYDROGENASE ;, mRNA
sequence.
ACCESSION AM152065
VERSION AM152065.1 GI:6199963
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 620)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.llnl.gov/bdrrp/image/image.html
Seq primer: -40UP from Glibco
High quality sequence stop: 416.
Location/Qualifiers
1. 620
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2623752"
/clone_lib="NCI CGAP Gas4"
/tissue_type="poorly differentiated adenocarcinoma with
signet ring cell features"
/lab_host="DH10B"
/note="Organ: stomach; Vector: pCMV-SPORT6; Site: 1: SalI;
Site: 2: NotI; Cloned unidirectionally. Primer: oligo dT.
Average insert size 1.69 kb. Life Technologies catalog #: 11549-011"

BASE COUNT 162 a 144 c 101 g 209 t 4 others
ORIGIN

Query Match 22.7%; Score 581; DB 111; Length 620;
Best Local Similarity 98.2%; Pred. No. 2.3e-90;
Matches 606; Conservative 0; Mismatches 9; Indels 2; Gaps 2;

QY 1404 gaaagcaggtgtagaaggtgatcatctcagaatgcagcatatcttccaaaag 1463
|||||
DB 620 GAAGCCCAAGTGTAGAAAGTTGATATCATGCGAAGTGCAGCATATTCATTTCCAAAA 561
|||||
QY 1464 ccaaaaag-tttactggcaacttgt-catgatgaataatcttaaaaagaagaaggaa 1521
|||||
DB 560 GCAAAAAGTTTACTGCAACTNCTGTCATGTGATGAATAATATCTTAAAGAGAGAA 501
|||||

```
QY 1522 tagaaatttgcgttctatgcaatlaaacacagtcattccttctgcaaccaagatttctc 1581
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 500 TAGANATTGTCAGCTTATGCAATTAACCAAGTCATCTTGGACCAAGATTTCTCT 441
QY 1582 tagatgaataaccagaagcagtttagcaagaagtgtgaatcgaactgtgtctgtccgaat 1641
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 440 TAGATGAAATACCCAGAGAGAGCTTAGCAAGAAAGTGAAATCACTGCTGCTGTCAGAAAT 381
QY 1642 tcaaaagaagaacacgtgcagctgcacacaaacacacgtctgtgagctgtggaagaacat 1701
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 380 TCANAAGAAAGAAAGTGCAGCTGCACCAACCAACACGCTTGTGAGCTGTGGAAGAAACAT 321
QY 1702 ttagaattgttaagaactctcactcagtgatgattgtttaaagccactcaagaactatc 1761
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 320 TTAGAATTGTTGAAGACTCTCTCAGATGATGATGTTGTTAAAGCCACTCAAGCAATCTATC 261
QY 1762 tgttgaactctccgttgaaagtgtgtgagcgtgtgttcttgaatcgaagaagaggtg 1821
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 260 TGTGTGAACCTCTCCGGTGAGATGCTGSCACGCTGTTCTTGATCTGAAAGCAAGGCTG 201
QY 1822 ggaatgcgatatgagaagcctctcagtcagcagcagatgtgtgatatgactactg 1881
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 200 GGAATGTCGATATGAGAGCCCTTCGATCAGGACATGTGTGATGATGATGATGATGATG 141
QY 1882 atgactgttaaaatgcttcaaggagaacaaacaaacaaatgcatcattcagga 1941
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 140 ATGACTTTTAAATAATGTTTTCAGGGAACTTAACCAACAAATGCTATCATGTCTCAGGA 81
QY 1942 aattgaagaattaaagttacatgagccctagcaatcaatctgaagaagcctaagaatcaga 2001
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 80 AATTGAAGATTAAAGTAACTAGCCCTTCAGCAATCAATTTGGAAGCTTAATGAATCAGA 21
QY 2002 tgaatgcagcagactgtga 2018
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 20 TGAATGCCAGACTGTGA 4

RESULT 8
BF699109 887 bp mRNA EST 22-DEC-2000
LOCUS 602126730F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4283671 5',
DEFINITION mRNA sequence.
ACCESSION BF699109
VERSION BF699109.1 GI:11984517
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 887)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1115 row: 1 column: 08
High quality sequence stop: 611.
Location/Qualifiers
1..887
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4283671"
/clone_lib="NIH_MGC_56"
/tissue_type="primative neuroectoderm"
/lab_host="DH10B (T1 phage-resistant)"
```

```
/note="Organ: brain; Vector: pDNR-LIB (Clontech); Site:1:
SfiI (ggcgccctcgcc); Site:2: SfiI (ggcgcttaagcc);
Double-stranded cDNA was prepared from cell line RNA. 5'
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-ATTCTAGAGCGGCGGCGCGGCGGATG-3' and 3' adaptor
sequence: 5'-ATCTAGAGCGGCGGCGGCGGCGGCGGATG-3' (30)BN-3'.
Average insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
BASE COUNT 251 a 185 c 227 g 224 t
ORIGIN
Query Match 22.3%; Score 571.4; DB 168; Length 887;
Best Local Similarity 95.5%; Pred. No. 1e-88; 26; Indels 3; Gaps 3;
Matches 620; Conservative 0; Mismatches 26; Indels 3; Gaps 3;
QY 1114 tgaatctgaatgaacgctgaacacacagagacacactcactgcatcctaagaatgattc 1173
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 12 TGGATCTGATGATGAGAGCTGAACACCAAGGACCACTACCTGCAATCTAAAGCATGTATTC 71
QY 1174 ctatttgaagaagcaagttgtctcatatcctaatactagtcaccactgaacctaa 1233
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 72 CTATTTGAAAAAGACCAAGTTGCTCATATCTCTCAATATCAGTCCACCACTGAACCTAA 131
QY 1234 atccagcttggttcaaacagcactgtgcttatccactcactgaatgattgtaagctatgt 1293
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 132 ATCCAGTTTGGTTCAACACGACACTGCTTATACCTTGTCTAAGTATGATGTATGTATGT 191
QY 1294 atgtgcttggaatgagcagaagaattaaagtgtgaatgagcagtcataatgagccta 1353
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 192 ATGCTGCTTGGAATGCGAGAAAGATTTAAAGTGGAATTCAGTCAATGATTAATGCCCCA 251
QY 1354 aaacagcacaacacactgctgctatgatatgctggaagagcactgtgtatcagaagccagt 1413
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 252 AANAGCCCATACACACTGCTGCTATGATATGCTGGAGAGACCTGCTATGCAAGCCCAT 311
QY 1414 gtagaaagttgatatacatctgacagatgcagcatalatccattccaaaagccaaaagtt 1473
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 312 GTAGAAAGTGTATCATCTGATGAGATGAGCATATTCATTTCCAAAAGCCAAAAGTT 371
QY 1474 ttaactggaacttgcattgatgataaataatcttaaaagaagaagaataagaataatttg 1533
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 372 TTAATGCAACTTTGTCATTTGATGAATAATCTTTAAAGAGAGAAATGAATAATTTTG 431
QY 1534 acgttatgcaattaaacacaggtcatcctcttgcaacacagattcctcttagatgaatacc 1593
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 432 ACCTTTATGCAATTAACCAAGGTCATCTTTGCAACCAAGTTTCTTATGATGAATACC 491
QY 1594 cagaagcagttgaagaag-aaagtgaatcaactgtgtctgttccagaatcagaagaag 1652
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 492 CAAAGCAGTATGACAGAAAGAGTGAATCACTGAGTG-TCACAAATTTCAAGAGAG 550
QY 1653 aaactgcagctgcaacc-aaacacagcttcctgagctgtggaagaacattagaattgt 1711
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 551 AAATCTCAGCTGCAACCAAAACACAGTTCTGAGCTGGAGAAACCAATTAATGTT 610
QY 1712 taagagctctcctcagtgatgattgtttaaagcactcaagaacatcat 1760
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 611 AGGACTCTCTCCGTCATTTGATGTGCTTAAGCCCTCAGCAACTATCTGT 659

RESULT 9
AW966207 541 bp mRNA EST 01-JUN-2000
LOCUS AW966207
DEFINITION EST178280 MAGE resequences, MAGI Homo sapiens cDNA, mRNA sequence.
ACCESSION AW966207
VERSION AW966207.1 GI:8156043
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
```

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt, I. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and Quackenbush, J.

TITLE Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray

JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 224

FEATURES
source Location/Qualifiers
1..541
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGI"
/note="Vector: pBluescriptSkm"

BASE COUNT 172 a 110 c 122 g 137 t
ORIGIN

Query Match 20.4%; Score 523.6; DB 122; Length 541;
Best Local Similarity 99.1%; Pred. No. 1.7e-80;
Matches 537; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 745 tctgtatctcgaagaatcagttaccccaacacgggagctgtgagagatgtacagtttca 804
DB 1 TCTGATCTACGAAGACATGATCCCAACACCGGAGCTGCGAGATGTACAGTTTAA 60

QY 805 tcacagttcaagccgtgtgcatgtgcaagctatgtcatgaaagcagcaagatgtgag 864
DB 61 TCACAGTTCAGACCCCTGGCATTGGCAAGCTATTTCATTGAAAGACGCAAGATGGAG 120

QY 865 caaatatgttatgtctgcaaaagccgcccagccacatccaaaactctagacaaatc 924
DB 121 CAAATATTTCTATTGCTGCAAAAGACCGCCACACATCCAAATCTTAGGCACATCT 180

QY 925 ataccgtctgcaagaatgtgaagcagctgtgaggaagccctgtccatgtatgttga 984
DB 181 ATACTGCTGCTGAAGAAATTTGAAGCAGTTGAGGAAAGCCCTTGCCATGTATTGTGATG 240

QY 985 tgaagatgaacagcagatcagtgctgcaagtgaagaagccatcaagaatgtgagaa 1044
DB 241 TGAAGATGAACAGCAGATCAGTGTGCAAGTGAAGAAACCATCAAGAAATTTGGAGGA 300

QY 1045 ttgatatctcgtttaaataatgcacgtgcatatgtttgacaaatcatgtgacacacta 1104
DB 301 TTGATATCTGTGTAATATATGCGAGTGCATTAGTTGACCAATATATGACACACCTA 360

QY 1105 ccaagagatgtgcatgtgatgaaagcgaacacgaagagacactacacttgcacaaag 1164
DB 361 CCAAGAGATGTGATCTGATGATGAACGTAACACACGAGGCACTCTGCTGATCTAAAG 420

QY 1165 catgtatccttatcttgaaaagaagaagtgtcatatctccaatcatcagctccccc 1224
DB 421 CATGTATCTCTATTATTGAAAAA-AGCAAAAGTTGCTCATATTTCATCAATATCAGCCAC 479

QY 1225 tgaacctaactcagtttgttcaaacacgacactgtgctataccatgtcctaagtatgta 1284
DB 480 TGAACCTAATTCAGTTTGGTTCAAAACAGCAGCTGTCTTAATCCATTGCTAAGTATGGA 539

QY 1285 tg 1286
DB 540 TG 541

RESULT 10

BF151259

LOCUS BF151259 675 bp mRNA EST 29-DEC-2000
DEFINITION uz14a11.y1 NCI-CGAP_Mam5 Mus musculus cDNA clone IMAGE:3669020 5'
similar to TR:09VB10 Q9VB10 CG5590 PROTEIN.; mRNA sequence.
ACCESSION BF151259
VERSION BF151259.1 GI:11032654

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 675)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/resources.shtml

MGI:1429788
Seq primer: -40RP from G1bco
High quality sequence stop: 434.

FEATURES
source Location/Qualifiers
1..675
/organism="Mus musculus"
/strain="C57/B6"
/db_xref="taxon:10090"
/clone_lib="IMAGE:3669020"
/clone_lib="NCI-CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SPORE6; Site:1: Salt;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Rodin Humphreys,
NIH"

BASE COUNT 195 a 140 c 171 g 168 t 1 others
ORIGIN

Query Match 20.3%; Score 518.8; DB 144; Length 675;
Best Local Similarity 85.5%; Pred. No. 1.2e-79;
Matches 577; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 747 tgatctcgaagaatcagtttaccacacccgagagctgtgagatgtacagttttatc 806
DB 1 TGAATCTGCGGGGTCTATGCTGCCCAACTGGAACCTGAGGATGACACAGTTTATC 60

QY 807 acaggtgcaagccgtgtgcatgtgcaagctatgtcatgaaagcagcaagatgtgagaa 866
DB 61 ACAGGTGCAAGCCGAGGCAATGGCAAGCAATTCCTTGAAGCCGCAAGATGAGACC 120

QY 867 aatattgttatgtctgcaaaagccgcccacatccaaactctagacaaatcat 926
DB 121 AATATTGTATTGCTGCGAAGACCAACCAACCCGAAACTCTCTCGGCACAATCTAC 180

QY 927 actgctgctggaagaatgtgaagcagttggaagaagccctgtccatgtatgttga 986
DB 181 ACGGCTGCTGAAGAAATTTAAGCAGCTGGAAGGACCGCTTGCTTGTGTGATGTCG 240

QY 987 agagatgaacagatcagctgtcagtggaagaagccatcaagaatgtgagaaatc 1046
DB 241 AGAGATGAACAGCAATTCACAGTGCAGTGAGAAAGCTGTGAGAAATTTGGAGCAATT 300

Oy	1047	gatatctcggtaaaatagccagtgccatttagttgcaccaatacatcttgcacaccactacc	1106
Db	301	cattatttttggtgaacaatgccagtcgtatttaccttgaccaccaaacgttgacactccgacc	360
Oy	1107	aagagaattgatcgtatgaacgtlgaacaccacgaaggaccactacccttgcattcaaaagca	1166
Db	361	aagacagtggacttatgatgtaatgtgaacaccacgaggcgaccctaccttgcattccaagca	420
Oy	1167	tgtatccttatttggaaaaagagcaagttgtctcatctatcctaataatagtccaccactg	1226
Db	421	tgtatttccctttttttaaanaagacnaagatgtcatatttctaattctcagcccattccctg	480
Oy	1227	aacctaaalccacagtttggtttccaacaaagcacctgtgtctatataccatttgtaaatgttatg	1286
Db	481	aaacctatfaccacctgtgtgttcaaacagacacgctgtgttatnaccattggccattatgatggcang	540
Oy	1287	tctatgtatgtccttgcgaatgcgacgaagaatttcaaagtgtgaattgcagtaatgatcata	1346
Db	541	tctatgtgtgtgtccttgggtgtgtgaatgtgaagatgtggaattgtgcaattgtgtgttta	600
Oy	1347	tgccttaaacacgacacatacacactcgtctgtatgataatgtctggagagacctgtgatcga	1406
Db	601	tgcgtttagacagctcatcttgcacactcgtctgtatgagatattcttggagagacttgcgtgtat	660
Oy	1407	agccagttgtaaaa	1421
Db	661	aattcattgtgagaaaa	675
RESULT 11			
Bf182478	815 bp	mRNA	EST 31-OCT-2000
LOCUS	601804245P1	NCL.CGAP_Mam5 Mus musculus cDNA clone IMAGE:4035179	5'
DEFINITION	mRNA sequence.		
ACCESSION	Bf182478		
VERSION	Bf182478.1	GI:11060621	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus.		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 815)		
AUTHORS	NIH-MGC http://mgc.ncl.nih.gov/.		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Tel.: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LHAM9308 row: j column: 12 High quality sequence stop: 680.		
FEATURES	Location/Qualifiers		
source	1..815		
	/organism="Mus musculus"		
	/strain="C57/B6"		
	/db_xref="taxon:10090"		
	/clone_image="IMAGE:4035179"		
	/clone_lib="NCL CGAP Mam5"		
	/tissue_type="tumor, gross tissue"		
	/dev_stage="7 months"		
	/lab_host="DH10B"		
	/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: salt; Site_2: NciI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Robin Humphreys, NIH."		
BASE COUNT	250 a 156 c 211 g 198 t		

ORIGIN

Query Match	20.1%;	Score 515.6;	DB 145;	Length 815;
Best Local Similarity	83.1%;	Pred. No. 4.1e-79;		
Matches 635;	Conservative 0;	Mismatches 124;	Indels 5;	Gaps 4;

Oy	903	cccaactcttaaggacacaatcctactgcgtcggaagaattgaaagcagtcttgaggagaag	962
Db	1	CCGAAATCTCTCGGCACAAATCTACAGCGCTGCTGAAGAANAATTGAAACCAGCTGGAGGACG	60
Oy	963	gcccttcgcatgtaactgtctga tgcTgaagatgbaacagacagtcagtgctgcagttggagaaa	1022
Db	61	GCTTCGCCCTGTGTGTTGATGTGAGAGATGAACAACCAATTAACAGTGTCACTGGAGAAA	120
Oy	1023	ggcctcaagaanaatttggagaatctgatctctgtttaatacaatgcccagtgccattgcttg	1082
Db	121	GCTTGAGAAATTTGGAGGAATGTGATTTTGGTAACAAATGCCAGTCTTTTAGCTTG	180
Oy	1083	accaatcacattggaacacacctaccagaagatatgactctatgataagctgaacaccaga	1142
Db	181	ACCAAACGCTTGGACACTCGCACCAGAGAGTSGACTTGTATGATGATGTGAACACACAG	240
Oy	1143	ggcaactaccttgacatctaagaacatgatatctcttatcttgaaaagagcaagtgtgccat	1202
Db	241	GGCACCTACCTTATCATCCAAGCATGTATTCCTTTTAAAAAAGCAAAATAGTGT	300
Oy	1203	atcccaaatcagtcacacacacgaaacctaacatccagttcttgctcaaacagacgtgct	1262
Db	301	ATTGTCAATCTCAACCCCAACCCCTTGAACTTAACCCACTGTGTGTAACAGCACTGTGCT	360
Oy	1263	tataccaatgctaaagatggtatgctctbatgatalgctcttggaatggcagaagaattcaa	1322
Db	361	TATACCACTTGGCAAAATATGGCATGTCTATGTCTGTCTGGGTGGCTGCAAAATTTAGA	420
Oy	1323	ggttaaatcagtcacatgatcatatgagcccaaaagacacatacacactgctctatgat	1382
Db	421	GGTAAATTCGACGTCATGCTTATGCGCTTGAAGAACGCCATTACACTGCTCTTATGGAT	480
Oy	1383	atgctggagagacctggtalcatgaaagccagtgtagaaaaattgatalcatacttgcaatgca	1442
Db	481	ATGCTGGAGAGATCTGGTGTGAAAMCAATGTAGAAAGTTGACATCAATTGCGAGCGT	540
Oy	1443	gcatatccactttccaaaagccaaaagtttacaggcaacttggcatltga--tgaaa	1500
Db	541	G-ATAATTCATTTT-CAAAAGCGAAAAAGTTTTACGTGCAACTTATTTATGAATGAAAA	598
Oy	1501	atatcttaaaagaagaagatatgaaaatcttgacgtctatgcaatlaaacaggatcatc	1560
Db	599	ATATCTTAAAAAAGAAGGAATCAACAAATTTGATCTTACGCAATTTTCACACAGTCAAT	658
Oy	1561	ccttgcaaccagatctctctcttgatgataatccccagaagcagtttagcaagaagtfgaat	1620
Db	659	CCCTGTTAGCGAATTTCTTGTAGATGACAGAACCAATCAGATATGGGGAGGAAAAAAT	718
Oy	1621	caactggctgctgttcacgaattcaagaagaagagaactgcagctg	1664
Db	719	CACATGAT-CTGTCCCGGAAGTTGGAGAAAGAAAGCGGCCGCG	761
RESULT	12		
LOCUS	AM044841		
DEFINITION	AM044841 693 bp mRNA EST 18-SEP-1999		
VERSION	unl4601.YI Sugeno mouse kidney mRna Mus musculus cDNA clone IMAGE:2192280 5' similar to WP_045B11.3 CE05427 ALCOHOL DEHYDROGENASE ; , mRNA sequence.		
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerogonathii; Muridae; Murineae; Mus		

Best Local Similarity 84.9%; Pred. No. 1.7e-77;
Matches 614; Conservative 0; Mismatches 103; Indels 6; Gaps 4;

QY 752 tgcgaagatcgtatccacacacccgagcgctgagcagatctacatctttatccag 811
1 tgcgaagatcgtatccacacacccgagcgctgagcagatctacatctttatccag 60
DB 1 tgcgaagatcgtatccacacacccgagcgctgagcagatctacatctttatccag 60
QY 812 tgcgaagcgttgcagatcgtacgaagctatctgcatgaaagcagaagatgagcaaat 871
1 tgcgaagcgttgcagatcgtacgaagctatctgcatgaaagcagaagatgagcaaat 1111
DB 61 tgcgaagcgttgcagatcgtacgaagctatctgcatgaaagcagaagatgagcaaat 120
QY 872 tgcgtatctgctgaagacccgcccagccacatctcaaaactctcagcacatctatctc 931
1 tgcgtatctgctgaagacccgcccagccacatctcaaaactctcagcacatctatctc 1111
DB 121 tgcgtatctgctgaagacccgcccagccacatctcaaaactctcagcacatctatctc 180
QY 932 tgcgtgaagaatctggaagcagcttgcgaagaaagccctgcacatctatctgcatgagaga 991
1 tgcgtgaagaatctggaagcagcttgcgaagaaagccctgcacatctatctgcatgagaga 1111
DB 181 tgcgtgaagaatctggaagcagcttgcgaagaaagccctgcacatctatctgcatgagaga 240
QY 992 tgaacacagcagatcagtcgctgcagctggaagaagccatcaagaatcttgagaagatgatat 1051
1 tgaacacagcagatcagtcgctgcagctggaagaagccatcaagaatcttgagaagatgatat 1111
DB 241 tgaacacagcagatcagtcgctgcagctggaagaagccatcaagaatcttgagaagatgatat 300
QY 1052 tctgtgtaaatgagtcagcagcagcattgattgacccaatctgagcacacacccaagag 1111
1 tctgtgtaaatgagtcagcagcagcattgattgacccaatctgagcacacacccaagag 1111
DB 301 tctgtgtaaatgagtcagcagcagcattgattgacccaatctgagcacacacccaagag 360
QY 1112 attggaatctgcatgatacgaatgaaacacagagcagcactctgcatctcaagcatgatat 1171
1 attggaatctgcatgatacgaatgaaacacagagcagcactctgcatctcaagcatgatat 1111
DB 361 attggaatctgcatgatacgaatgaaacacagagcagcactctgcatctcaagcatgatat 419
QY 1172 tccattatctggaagaagcagaagctgctcatatctcctaataatcagtcacacacacacac 1231
1 tccattatctggaagaagcagaagctgctcatatctcctaataatcagtcacacacacacac 1111
DB 420 tccattatctggaagaagcagaagctgctcatatctcctaataatcagtcacacacacacac 479
QY 1232 aatccaagcttgctgaac 1291
1 aatccaagcttgctgaac 1111
DB 480 aatccaagcttgctgaac 539
QY 1292 gtaatgcttgcgtgaagcagcagaagcattcaaaagctgaagcattgacatgcatatgagc 1351
1 gtaatgcttgcgtgaagcagcagaagcattcaaaagctgaagcattgacatgcatatgagc 1111
DB 540 gtaatgcttgcgtgaagcagcagaagcattcaaaagctgaagcattgacatgcatatgagc 597
QY 1352 taaacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1411
1 taaacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1111
DB 598 -TAGAAGACACATTCACACTGCTGCTATGCTATGCTGAGAGGATCTGCTGAGAAACCAA 656
QY 1412 gtcgtgaagaagctgcatatctgcatgagcagcagcagcagcagcagcagcagcagcagc 1471
1 gtcgtgaagaagctgcatatctgcatgagcagcagcagcagcagcagcagcagcagcagc 1111
DB 657 gtcgtgaagaagctgcatatctgcatgagcagcagcagcagcagcagcagcagcagcagc 714
QY 1472 tct 1474
1 tct 1474
DB 715 TTT 717

RESULT 14
BE307752 693 bp mRNA EST 26-OCT-2000
LOCUS 60109805221 NCI_CGAP_Mam5 Mus musculus CDNA IMAGE:3496741 5'
DEFINITION mRNA sequence.
ACCESSION BE307752.1 GI:9163339
VERSION BE307752
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1998)
JOURNAL Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM8549 row: g column: 14
High quality sequence stop: 654.

FEATURES
source location/Qualifiers
1..693
/organism="Mus musculus"
/strain="C57/B6"
/db_xref="taxon:10090"
/clone="IMAGE:3496741"
/clone_1ib="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-Sport6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

BASE COUNT 202 a 153 c 173 g 165 t
ORIGIN

Query Match 19.2%; Score 492.8; DB 165; Length 693;
Best Local Similarity 85.7%; Pred. No. 3.5e-75;
Matches 595; Conservative 0; Mismatches 92; Indels 7; Gaps 4;

QY 744 ccttgatcctcaagaagctgcttaccacacacgagcagcagcagcagcagcagcagcagcagc 803
1 ccttgatcctcaagaagctgcttaccacacacgagcagcagcagcagcagcagcagcagcagc 1111
DB 1 ccttgatcctcaagaagctgcttaccacacacgagcagcagcagcagcagcagcagcagcagc 56
QY 804 atccacagctgcaagccgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 863
1 atccacagctgcaagccgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1111
DB 57 atccacagctgcaagccgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 116
QY 864 gcaaatatctgcatatctgctgcaagaagcagcagcagcagcagcagcagcagcagcagcagc 923
1 gcaaatatctgcatatctgctgcaagaagcagcagcagcagcagcagcagcagcagcagcagc 1111
DB 117 GCCAATATTGCTATGCTGCTGCAAGACACCCCAAAACACCCCAACTCCTCGCACAAATC 176
QY 924 taaactgctgcgaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 983
1 taaactgctgcgaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1111
DB 177 TACACGCGCTGCTGCAAAATTAAGACGCTGAGAGGACGCGCTTGCTTGTGTAAT 236
QY 984 gtcgaagatgaaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1043
1 gtcgaagatgaaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1111
DB 237 GTGAGAGATGAACAGCAATCAACACTGAGAGGAGAAAGCTGTGGAANAATTTGAGAGA 296
QY 1044 attgataatctgtaataatgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1103
1 attgataatctgtaataatgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1111
DB 297 ATTGATATTGCTGCAACATGCCACTGCTATTAGCTTACCAACACGTTGACACACTCGG 356
QY 1104 accaagagatctgcatctgtaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1163
1 accaagagatctgcatctgtaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1111
DB 357 ACCAAGAGAGTGAAGCTTGATGAATGGAACACACGAGGAGCACTCACTTCAATCCAAA 416
QY 1164 gcatgtatctcctatctgtaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1223
1 gcatgtatctcctatctgtaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1111
DB 417 GCATGATTCCTTTTAAAGAGCAAAAGTAGATCATATCTCAATCTGAGCCACCC 476
QY 1224 ctgaacctaataacagcttggcttcaaacagcagcagcagcagcagcagcagcagcagcagcagc 1283
1 ctgaacctaataacagcttggcttcaaacagcagcagcagcagcagcagcagcagcagcagcagc 1111
DB 477 CTGAACCTTAACCACTGCTGTTCAAAAGCACTGCTTAAACCAATTCGCAAAATATGCG 536
QY 1284 atgtctatct-gtatgcttgcgtgaagcagcagcagcagcagcagcagcagcagcagcagcagc 1342
1 atgtctatct-gtatgcttgcgtgaagcagcagcagcagcagcagcagcagcagcagcagcagc 1111
DB 537 ATGTCTATGTGCTGCTTGGAGTGGCTGAAGAAATTTAAGAGGTGAATTTGAGTCAATGCG 596

QY 1343 attatgacctaaacagccatcacactgctc-atgagatgctggaagacctgta 1401
|||||
DB 597 CTTATGGGCT-AAACAGCCATTACACCTGCTGTAATGATGCGAGATCTGTCG 655
QY 1402 tcgaagccagctgtagaagatgatatcatctgc 1435
|||||
DB 656 TTGACACCACTGTAGAACACTTGCATTCATTGC 689

RESULT 15
BG245008 618 bp mRNA EST 13-FEB-2001
LOCUS BG245008 602358280P1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4486915 5',
DEFINITION mRNA sequence.
ACCESSION BG245008
VERSION BG245008.1 GI:12754823
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 618)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM10330 row: P column: 20
High quality sequence stop: 618.

FEATURES
source location/Qualifiers
1..618
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4486915"
/clone_1id="NCI_CGAP_Mam1"
/tissue_type="tumor, biopsy sample"
/dev_stage="3 months, virgin"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SPORT6; Site:1: Salt;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT 183 a 134 c 156 g 145 t
ORIGIN

Query Match 19.0%; Score 485.4; DB 174; Length 618;
Best Local Similarity 87.6%; Pred. No. 6.5e-74;
Matches 542; Conservative 0; Mismatches 76; Indels 1; Gaps 1;

QY 773 caccggaggagcctgagcagatgttattcaagatgcaagccgctgcatggcaa 832
|||||
DB 1 CACTGGGAAGCTAGCGGAGTGCACAGTTTTCACAGGTGCAGCGCATTTGGCA 60

QY 833 agctattgcatgaaagcagcaagagatgagcaaatatgttatgtctgcaagaccgc 892
|||||
DB 61 AGCAATTGGCTTGAAAGCGCAAGAGATGAGCCATATTGTCATTGCTGCGAAGCAC 120

QY 893 ccagccacatcccaaacctctagcacaactctatactgctgctggaagaattggaagcag 952
|||||
DB 121 CCAAAAGCACCCGAAACTCTCGGCACATCTACAGGCTGCTGAAGAAATTGAAGCAC 180

QY 953 tggaggaagagccttgcatgtatgttgatgagatgaaacagcagatcagtgctgc 1012

DB 181 TGGAGGAGAGCGGCTTGTCCCTTGT-GTCTTATGTGAGAGATGAAACAGCAATCAACAGTGC 239
|||||
QY 1013 agtggagaagccatcaagaanaattggaggaattgatalctctgtaataatgacagtcg 1072
|||||
DB 240 AGTGGAGAAAGCTGTCGACAAATTTGAGCAATTTGATTTATTTGTCGACAAATGCCAGTGC 299

QY 1073 catgtgttgacccaatacatctggacacacctcccaagagatgtggatctgatatgaagt 1132
|||||
DB 300 TATTAGCTTGACCAACACGCTTGACACCTCCGACAGAGAGAGTGAAGTATGATGAATGT 359

QY 1133 gaacacagagcagaccttaccatctgcatctcaagcatgtatcttatttgaagaagacaa 1192
|||||
DB 360 GAACACGAGGGGACCTTACCTTACATCCAAAGCATGTATTCTTTTAAAAAGACGA 419

QY 1193 agtgcatactcctcaataatacagtcacacactgaaacctaaatcagatttgltcaaca 1252
|||||
DB 420 AGTAGGTCAATATTTCATCAATCTCAGCCACCCTGAACCTAAACCCACTGTGTTCAACA 479

QY 1253 gcaactgtcttataccatgtcgaagtatgtctatgtatgtgctggaatggcaga 1312
|||||
DB 480 GCACTGTGCTTATACCATTTGCCAAATATGGCATGTCTATGTGCTTGGAGTGGCTGA 539

QY 1313 agaattlaaggatgaattgcagtcacatgcatatgacctaaacagccatacacactgc 1372
|||||
DB 540 AGAATCTAGAGTGGAATTCAGAGTCAATGCTTATGCGCTTAGAACAGCATTCACACTGC 599

QY 1373 tgcatagatatgctggga 1391
|||||
DB 600 TCTATGATATGCTGGGA 618

Search completed: June 12, 2001, 10:40:35
Job time: 4553 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 12, 2001, 09:59:37 ; Search time 89.72 Seconds
(without alignments)
4985.922 Million cell updates/sec

Title: US-09-464-039-8

Perfect score: 2561

Sequence: 1 agcgacgaagtatgcaaacga.....gsgmggraswmwawrrammc 2561

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6C.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/6C.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	556.2	21.7	4951	2	US-08-752-307B-1
2	400	15.6	633	1	US-08-644-664B-6
3	400	15.6	633	2	US-08-761-277A-6
4	259.4	10.1	9045	3	US-09-121-321-1
5	249.4	9.7	720	1	US-08-217-210B-3
6	181.4	7.1	422	1	US-07-920-519-33
7	181.4	7.1	422	1	US-08-314-586-33
8	181.4	7.1	422	1	US-08-314-586-33
9	181.4	7.1	422	2	US-07-927-391-14
10	167.6	6.5	7287	2	US-08-659-206A-1
11	166	6.5	6253	3	US-08-893-327-15
12	166	6.5	6280	3	US-08-893-327-17
13	166	6.5	6280	3	US-08-893-327-19
14	166	6.5	6295	2	US-08-659-206A-4
15	166	6.5	7852	4	US-08-836-022A-2
16	166	6.5	7852	4	US-09-427-048A-2
17	166	6.5	7892	2	US-07-916-098A-40
18	166	6.5	7897	4	US-08-836-022A-1
19	166	6.5	7897	4	US-09-427-048A-1
20	166	6.5	8236	5	US-08-461-837-1
21	166	6.5	8236	5	US-08-973-223-1
22	166	6.5	8236	5	PCT-US93-09495-1
23	166	6.5	8309	1	US-08-462-014-1
24	166	6.5	10398	2	US-08-331-384-1
25	166	6.5	10398	2	US-08-708-188-1
26	166	6.5	10398	2	US-08-836-087-1
27	166	6.5	19307	3	US-08-836-022A-10

C 28	166	6.5	19307	4	US-09-427-048A-10	Sequence 10, Appl
C 29	163.6	6.4	7286	4	US-09-331-581-3	Sequence 3, Appl
C 30	163.6	6.4	7938	4	US-09-331-581-14	Sequence 14, Appl
C 31	163	6.4	5141	1	US-08-286-305A-9	Sequence 9, Appl
C 32	163	6.4	5141	2	US-08-441-104A-9	Sequence 9, Appl
C 33	163	6.4	5141	2	US-08-440-816A-9	Sequence 9, Appl
C 34	163	6.4	6557	5	PCT-US93-09576-3	Sequence 3, Appl
C 35	163	6.4	6557	5	PCT-US93-09576-3	Sequence 3, Appl
C 36	163	6.4	8120	3	US-09-027-449-68	Sequence 68, Appl
C 37	163	6.4	8120	4	US-09-026-985-68	Sequence 68, Appl
C 38	161.8	6.3	10596	1	US-07-884-811-15	Sequence 15, Appl
C 39	161.8	6.3	10596	1	US-07-885-971-15	Sequence 15, Appl
C 40	161.8	6.3	10596	1	US-08-087-783A-15	Sequence 15, Appl
C 41	161.8	6.3	10596	1	US-08-194-088B-15	Sequence 15, Appl
C 42	161.8	6.3	10596	2	US-08-194-087-15	Sequence 15, Appl
C 43	161.8	6.3	10596	5	PCT-US93-04648-15	Sequence 15, Appl
C 44	161.4	6.3	10785	3	US-08-444-644-27	Sequence 27, Appl
C 45	161.4	6.3	10844	3	US-08-444-644-41	Sequence 41, Appl

ALIGNMENTS

RESULT 1
US-08-752-307B-1
; Sequence 1, Application US/08752307B
; Patent No. 5952171
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Geating, David P.
; APPLICANT: Levinson, Douglas A.
; TITLE OF INVENTION: METHOD FOR IDENTIFYING GENES
; TITLE OF INVENTION: ENCODING NOVEL SECRETED OR MEMBRANE-ASSOCIATED PROTEIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; City: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/752,307B
; FILING DATE: 19-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: McKieJohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; TELEPHONE/DOCKET NUMBER: 09404/020001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4951 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-752-307B-1
Query Match 21.7%; Score 556.2; DB 2; Length 4951;
Best local similarity 96.2%; Pred. No. 3.7e-105;
Matches 656; Conservative 0; Mismatches 14; Indels 12; Gaps 8;

```

QY 1 aggcagaagtaatgcaaaagatgcatctcaaatatgtaacgaacatagtcggccct 60
DB 131 AGCGAAGATATGCAAAAGATGATCTC-AATTAGTCAGC-AACCATATATCC-GCCCT 187
QY 61 aactcgcacatcccgccctcaactccgncagagtcgcggccatctccggccatgac 120
DB 188 AACTCGCGCCATCCGCCCTCACTCCGCC--AGTTCGCCCATCTCCGCCCATGAC 245
QY 121 tgactaatcttttattatgatcaga-gccgagggccgctcgccctgagctatccag 179
DB 246 TGACTAATTTTATTTATATGACAGGCCGAGGCCCTCGGCTCTAGCTATTCAG 305
QY 180 aagtagtgaagagctcttttggagagctaggcttttgcacaaagctcct-cgataag 238
DB 306 AAGTAGTGAGAGGCTTTTGGAGGCTTAGGCTTTGCAAAAGCTCTCGATGAGAG 365
QY 239 ggcgcacatctctctcaacgcgcgcgcgcctcaactagtaggcgcacatccagcgtt 298
DB 366 GCGTCGACATCTCTCTTACGCGCGCGCCCTCACTGAGCGCGCATGACGCGGTT 425
QY 299 ggcgcacatctctctcaacgcgcgcgcgcctcaactagtaggcgcacatccagcgtt 358
DB 426 GAGTCGCGCTTCCGCGCTCCGCGCTGAGTCTCTGATGATGATGATGATGATGAT 485
QY 359 aagttaaagctcagagtcagagccgagccttgcgcgcgcctctggagcctacatag 418
DB 486 AAGTTAAAGCTCAGAGTCAGAGCGGCGCTTGTCCGCGGCTCCCTTGAGGCTACCTAG 545
QY 419 actcagcgcgcctcgaagcgttgcgcgcgcgcctgctcaactcaactgcttcttctc 478
DB 546 ACTCAGCGCGCTCAGAGCTTGTCCGCGCTGAGTCTCTGATGATGATGATGATGAT 605
QY 479 agttctcgtctcgcgcgcgtacagatccaagctctgaaacacagaaagcttaactgta 538
DB 606 -CTTTCTCTTCTGCGCGCTTACAGATCCAACTCTGAAAAACGAAAGTTAACTGTA 664
QY 539 agtttagctcttctcttcttcttcttcttcttcttcttcttcttcttcttcttct 594
DB 665 AGTTAGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 724
QY 595 aagaactctcctcagagtcagagtcgttcttcttcttcttcttcttcttcttcttct 654
DB 725 AAGAAGCTCTCTCAGAGTCAGAGTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 784
QY 655 tgctctaaagctcgcgaatc 676
DB 785 TGCTCTAAAGCTCGCGAATTC 806

```

```

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: GENIOTPE-00912
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 633 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-644-664B-6

Query Match      15.6%; Score 400; DB 1; Length 633;
Best Local Similarity 96.1%; Pred. No. 1.7e-73;
Matches 487; Conservative 0; Mismatches 6; Indels 14; Gaps 7;

QY 1 aggcagaagtaatgcaaaagatgcatctcaaatatgtaacgaacatagtcggccct 60
DB 131 AGCGAAGATATGCAAAAGATGATCTC-AATTAGTCAGC-AACCATATATCC-GCCCT 187
QY 61 aactcgcacatcccgccctcaactccgncagagtcgcggccatctccggccatgac 120
DB 188 AACTCGCGCCATCCGCCCTCACTCCGCC--AGTTCGCCCATCTCCGCCCATGAC 245
QY 121 tgactaatcttttattatgatcaga-gccgagggccgctcgccctgagctatccag 179
DB 246 TGACTAATTTTATTTATATGACAGGCCGAGGCCCTCGGCTCTGAGCTATTCAG 305
QY 180 aagtagtgaagagctcttttggagagctaggcttttgcacaaagctcct-cgataag 238
DB 306 AAGTAGTGAGAGGCTTTTGGAGGCTTAGGCTTTGCAAAAGCTCTCTGCA----- 358
QY 240 gtcgcacatctcctcgaagtcagcgcgcgcgcctcaactagtaggcgcacatccagcgtt 299
DB 359 GCTCGCATCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 418
QY 300 agtcgcgtctcgcgcgcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 359
DB 419 ACTCGGCTTCTCCGCGCTCCGCGCTGAGTCTCTGAACTGCTCCGCGCTCTAGGTA 478
QY 360 agttaaagctcagagtcagagccgagccttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 419
DB 479 AGTTAAGCTCAGAGTCAGAGCCGAGCCCTTGTCCGCGCTCTCTGAGGCTTACTAGA 538
QY 420 ctcagcgcgcctcgaagtcgttgcctgaaccctgcttgcctcaactcagcttcttctca 479
DB 539 CTCAGCGCGCTCTCCAGCGCTTGTGCTGAGCCCTGCTTGTGCTGAGCTTCTTCTTCT 597
QY 480 gttctcgtctcgcgcgcgttcaagatc 506
DB 598 GTTCTCTGTTGCGCGCTTACAGATC 624

RESULT 3
US-08-761-277A-6
Sequence 6, Application US/08761277A
Patent No. 5972334
GENERAL INFORMATION:
APPLICANT: Denney Jr., Dan W.
TITLE OF INVENTION: Vaccines For Treatment Of Lymphoma And
NUMBER OF SEQUENCES: 80
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California

```


Db 472 GTAAGTTAAAGCTCAGAGTCAGACCGGGCCCTTGTCCGGCGCTCCCTTGAGAGGCTACT 531
Qy 417 agactcagcgcgcctcctcagcgccttgctgacccctgctgctcaactcctacgtcttgc 476
|||||
Db 532 AGACTCAGCGCGCTCTCAGACGCTTGCTGACCTGCTGCTGCTCAACTACTAGCTTTGTT 591
Qy 477 tcaatttcgtctctgcgcgcttcacagatccaa 509
|||||
Db 592 TC-GTTTCTGTTCTCTGCGCCCTTACAGATCGAA 623
RESULT 5
US-08-217-210B-3
Sequence 3, Application US/08217210B
Patent No. 5580761
GENERAL INFORMATION:
APPLICANT: Wilson Greatbatch and John C. Sanford
TITLE OF INVENTION: Method of Confering Resistance to
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hodgson, Russ, Andrews, Woods & Goodyear
STREET: 1800 One Met Plaza
City: Buffalo
STATE: New York
COUNTRY: United States
ZIP: 14203-2391
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb storage
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS/ Microsoft Windows 3.1
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/217,210B
FILING DATE: 23/3/94
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/739,718
FILING DATE: 29/7/91
APPLICATION NUMBER: 07/156,188
FILING DATE: 16/2/88
ATTORNEY/AGENT INFORMATION:
NAME: Nelson, M. Bud
REGISTRATION NUMBER: 35,300
REFERENCE/DOCKET NUMBER: 06982,0006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 856-4000
TELEFAX: (716) 849-0349
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 720 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: RNA
IMMEDIATE SOURCE: genomic
ORIGINAL SOURCE:
ORGANISM: HTLV-1
CELL TYPE: virus
FEATURE:
OTHER INFORMATION: contains LTR region
US-08-217-210B-3
Query Match 9.7%; Score 249.4; DB 1; Length 720;
Best Local Similarity 71.6%; Pred. No. 1.3e-42;
Matches 194; Conservative 70; Mismatches 6; Indels 1; Gaps 1;
Qy 239 ggcctcgcctctctcctcagcgcgcgcgcctcactcctgagcgcgcacatcagcggtt 298
|||||
Db 1 GGCCTCGCAUCCUCUCACGCGCGCCGCCACCCUACCGAGGCGCCUCCAUCCACGCCGCAU 60
Qy 299 gagctgagctctgcgcgcctcgcgcgttgctgctcctgactcagctcgcgcgtctggt 358
|||||

Db 61 GAGCGCGUUCUGCCGCCUCCGCGGUGGUGCCUGACUGCGUCCGCCGUGUAGU 120
Qy 359 aagttaaagctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 418
|||||
Db 121 AAGUUAAGUCUGACGUGAGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
Qy 419 actcagcgcgcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 478
|||||
Db 181 ACUCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
Qy 479 agttctgctctgcgcgcgttcacagatccaa 509
|||||
Db 241 -GUUUUGUUCUGCGCGGUGACAGUCCAA 270
RESULT 6
US-07-920-519-33
Sequence 33, Application US/07920519
Patent No. 5382518
GENERAL INFORMATION:
APPLICANT: CAPUT, DANIEL
APPLICANT: FERRARA, PASCUAL
APPLICANT: GUILLEMOT, JEAN-CLAUDE
APPLICANT: KAGHAD, MOURAD
APPLICANT: LEGOUX, RICHARD
APPLICANT: LOISON, GERARD
APPLICANT: LABRE, ELIZABETH
APPLICANT: LUPKER, JOHANNES
APPLICANT: LEPLATOIS, PASCUAL
APPLICANT: SALOME, MARK
TITLE OF INVENTION: URATE OXIDASE ACTIVITY PROTEIN,
RECOMBINANT GENE CODING THEREFOR, EXPRESSION VECTOR,
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
City: Alexandria
STATE: Virginia
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/920,519
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/659,408
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16781/276 BDCL
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 422 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
IMMEDIATE SOURCE:
CLONE: plasmid pSEL "site binding to HindIII"
US-07-920-519-33

Query Match 7.1%, Score 181.4; DB 1; Length 422;
Best Local Similarity 85.6%; Pred. No. 9.3e-29;
Matches 237; Conservative 0; Mismatches 36; Indels 4; Gaps 3;

OY 239 ggtcgcacatctctcttaacgagccgagccgagcctgagagccgacatccagcggtt 298
|||||
DB 5 ggctggcattctctcttaacgagccgagccgagcctgagagccgacatccagcggtt 63
OY 299 gagtcgcttctgcgcgctccgcgctgtgtgctcctcgaactgctgcgcgctagt 358
|||||
DB 64 GAGTGGCGTTCTGCGCGCTCCGCGCTGTGTGCTCTCGAAGCTGCGCTCGCGCTAGGT 123
OY 359 aagtttaagctcaagtcgagacgagcgcttctgcgagcgctccttggagcctactag 418
|||||
DB 124 AGGCTCCCAAG--GGAGCCGAGCAAGGCCCGGCTCTGACCTGAGCTCTAACTTACTAG 181
OY 419 actagccgagctctcaagcttgcctgacgctgtgctgaactctagcttctgttct 478
|||||
DB 182 ACTAGCGCGCTCTCAAGCTTGTGCTTGCCTGACCTGCTCTCACTCTAGCTTTGTTTC 241
OY 479 agttctgtctgcgcgcttcaagatccaagctctg 515
|||||
DB 242 -GTTTCTGTCTGCGCGCTTACACTTCAAGGTATG 277

RESULT 7
US-08-314-586-33
Sequence 33, Application US/08314586
Patent No. 5541098

GENERAL INFORMATION:

APPLICANT: CAPUT, DANIEL
APPLICANT: FERRARA, PASCUAL
APPLICANT: GUILLEMOT, JEAN-CLAUDE
APPLICANT: KAGHAD, MOURAD
APPLICANT: LEGOUX, RICHARD
APPLICANT: LOISON, GERARD
APPLICANT: LARRE, ELIZABETH
APPLICANT: LUPKER, JOHANNES
APPLICANT: LEPLATOIS, PASCUAL
APPLICANT: SALOME, MARK
APPLICANT: LAURENT, PATRICK
TITLE OF INVENTION: URATE OXIDASE ACTIVITY PROTEIN,
TITLE OF INVENTION: RECOMBINANT GENE CODING THEREFOR, EXPRESSION VECTOR,
TITLE OF INVENTION: MICROORGANISMS AND TRANSFORMED CELLS
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/314,586
FILING DATE: 28-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/659,408
FILING DATE: 25-APR-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16781/509/BDL
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109

TELEX: 899149
INFORMATION FOR SEQ ID NO: 33:

SEQUENCE CHARACTERISTICS:
LENGTH: 422 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
IMMEDIATE SOURCE:
CLONE: Plasmid pSEL "site binding to HindIII"
US-08-314-586-33
CLONE: fragment

Query Match 7.1%, Score 181.4; DB 1; Length 422;
Best Local Similarity 85.6%; Pred. No. 9.3e-29;
Matches 237; Conservative 0; Mismatches 36; Indels 4; Gaps 3;

OY 239 ggtcgcacatctctcttaacgagccgagccgagcctgagagccgacatccagcggtt 298
|||||
DB 5 ggctggcattctctcttaacgagccgagccgagcctgagagccgacatccagcggtt 63
OY 299 gagtcgcttctgcgcgctccgcgctgtgtgctcctcgaactgctgcgcgctagt 358
|||||
DB 64 GAGTGGCGTTCTGCGCGCTCCGCGCTGTGTGCTCTCGAAGCTGCGCTCGCGCTAGGT 123
OY 359 aagtttaagctcaagtcgagacgagcgcttctgcgagcgctccttggagcctactag 418
|||||
DB 124 AGGCTCCCAAG--GGAGCCGAGCAAGGCCCGGCTCTGACCTGAGCTCTAACTTACTAG 181
OY 419 actagccgagctctcaagcttgcctgacgctgtgctgaactctagcttctgttct 478
|||||
DB 182 ACTAGCGCGCTCTCAAGCTTGTGCTTGCCTGACCTGCTCTCACTCTAGCTTTGTTTC 241
OY 479 agttctgtctgcgcgcttcaagatccaagctctg 515
|||||
DB 242 -GTTTCTGTCTGCGCGCTTACACTTCAAGGTATG 277

RESULT 8

US-08-371-121-13
Sequence 13, Application US/08371121
Patent No. 5652123

GENERAL INFORMATION:

APPLICANT: CAPUT, Daniel
APPLICANT: FERRARA, Pascual
APPLICANT: GUILLEMOT, Jean-Claude
APPLICANT: KAGHAD, Mourad
APPLICANT: LARRE, ELIZABETH
APPLICANT: LUPKER, JOHANNES
APPLICANT: LEPLATOIS, PASCUAL
APPLICANT: LABIT-LE BOUTELLIER, Christine
APPLICANT: MAGAZIN, Marilyn
TITLE OF INVENTION: Protein having a cytokine type
TITLE OF INVENTION: activity, recombinant DNA coding for this protein,
TITLE OF INVENTION: transformed cells and microorganisms.
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/371,121
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:


```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/659,206A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 110229.91144
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7287 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Plasmid DNA"
IMMEDIATE SOURCE:
CLONE: pMRG3169
FEATURE:
NAME/KEY: Promoter
LOCATION: 1..628
FEATURE:
NAME/KEY: IDNA
LOCATION: 629..810
FEATURE:
NAME/KEY: CDS
LOCATION: 1010(953..1258, 1332..1673)
OTHER INFORMATION: /product= "p35 gene product"
FEATURE:
NAME/KEY: polyA_site
LOCATION: 1797..2024
FEATURE:
NAME/KEY: Promoter
LOCATION: 2110..2737
FEATURE:
NAME/KEY: IDNA
LOCATION: 2738..2919
FEATURE:
NAME/KEY: CDS
LOCATION: 2983..3990
OTHER INFORMATION: /product= "p40 gene product"
FEATURE:
NAME/KEY: polyA_site
LOCATION: 4075..4306
US-08-659-206A-1

Query Match
Best Local Similarity 97.7%; Pred. No. 1.9e-25;
Matches 170; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 505 tccaagctctgaaaaaacgaagaagtaactggttaagtttagctcttcttcttatttc 564
DB 630 TCCAGGAAGCTGAAAAACGAAAGTTAACTGTAAGTTAGCTTTTGTCTTTATTTC 689
QY 555 aggtccggatccggtgtgtgtgtgcaaatcaaaagaactgctcctcagtggaattgcctt 624
DB 690 AGGTCCGGATCCGGTGTGTGTGCAAAATCAAGAACTGCTCTCAAGTGTGATGCTCTT 749
QY 625 actctagaagctgtacggaagtgtaactctgctcttaaaagctgcggaattcta 678
DB 750 ACTTCTAGGCGCTGTACGGAAGTGTACTTCTGCTCTTAAAAAGCTGCGGAATTGTA 803

RESULT 11
US-08-893-327-15
: Sequence 15, Application US/08893327
```

```
Patent No. 6020192
GENERAL INFORMATION:
APPLICANT: Zolotukhin, Sergei
APPLICANT: Hauswirth, William W.
APPLICANT: Muzyczka, Nicholas
TITLE OF INVENTION: Humanized Green Fluorescent Protein
TITLE OF INVENTION: Genes and Methods
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESS: Arnold, White & Durkee
STREET: P. O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/893,327
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/588,201
FILING DATE: 18-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kitchell, Barbara S.
REGISTRATION NUMBER: 33,928
REFERENCE/DOCKET NUMBER: UFLA:062\KIT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 6253 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 988..1701
US-08-893-327-15

Query Match
Best Local Similarity 97.1%; Pred. No. 3.7e-25;
Matches 169; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 505 tccaagctctgaaaaaacgaagaagtaactggttaagtttagctcttcttcttatttc 564
DB 796 TCGAGCAACTGAAAAACGAAAGTTAACTGTAAGTTAGCTTTTGTCTTTATTTC 855
QY 565 aggtccggatccggtgtgtgtgtgcaaatcaaaagaactgctcctcagtggaattgcctt 624
DB 856 AGGTCCGGATCCGGTGTGTGTGCAAAATCAAGAACTGCTCTCAAGTGTGATGCTCTT 915
QY 625 actctagaagctgtacggaagtgtaactctgctcttaaaagctgcggaattcta 678
DB 916 ACTTCTAGGCGCTGTACGGAAGTGTACTTCTGCTCTTAAAAAGCTGCGGAATTGTA 969

RESULT 12
US-08-893-327-17
: Sequence 17, Application US/08893327
: Patent No. 6020192
: GENERAL INFORMATION:
: APPLICANT: Zolotukhin, Sergei
: APPLICANT: Hauswirth, William W.
: APPLICANT: Muzyczka, Nicholas
: TITLE OF INVENTION: Humanized Green Fluorescent Protein
: TITLE OF INVENTION: Genes and Methods
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 12, 2001, 09:27:17 ; Search time 3233 6 Seconds

(without alignments)
11680.253 Million cell updates/sec

Title: US-09-464-039-8

Perfect score: 2561

Sequence: 1 aggcagagatgtgcaagca.....gsgmgaswmwawrramnc 2561

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_om:*
- 8: gb_ov:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_p11:*
- 13: gb_p12:*
- 14: gb_p13:*
- 15: gb_p14:*
- 16: em_ba1:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_higo_hum:*
- 20: em_higo_inv:*
- 21: em_higo_rod:*
- 22: em_hig_hum1:*
- 23: em_hig_hum2:*
- 24: em_hig_hum3:*
- 25: em_hig_hum4:*
- 26: em_hig_hum5:*
- 27: em_hig_hum6:*
- 28: em_hig_hum7:*
- 29: em_hig_hum8:*
- 30: em_hig_inv1:*
- 31: em_hig_inv2:*
- 32: em_hig_other:*
- 33: em_hig_rod:*
- 34: em_hum1:*
- 35: em_hum2:*
- 36: em_hum3:*
- 37: em_hum4:*
- 38: em_hum5:*
- 39: em_hum6:*
- 40: em_hum7:*
- 41: em_in:*
- 42: em_om:*
- 43: em_or:*

- 44: em_ov:*
- 45: em_pat:*
- 46: em_ph:*
- 47: em_p1:*
- 48: em_rod:*
- 49: em_sts:*
- 50: em_sy:*
- 51: em_un:*
- 52: em_v1:*
- 53: gb_sts1:*
- 54: gb_sts2:*
- 55: gb_sts3:*
- 56: gb_sy:*
- 57: gb_un:*
- 58: gb_v11:*
- 59: gb_v12:*
- 60: gb_hig1:*
- 61: gb_hig2:*
- 62: gb_hig3:*
- 63: gb_hig4:*
- 64: gb_hig5:*
- 65: gb_hig6:*
- 66: gb_hig7:*
- 67: gb_hig8:*
- 68: gb_hig9:*
- 69: gb_hig10:*
- 70: gb_hig11:*
- 71: gb_hig12:*
- 72: gb_hig13:*
- 73: gb_hig14:*
- 74: gb_hig15:*
- 75: gb_hig16:*
- 76: gb_hig17:*
- 77: gb_hig18:*
- 78: gb_hig19:*
- 79: gb_hig20:*
- 80: gb_hig21:*
- 81: gb_hig22:*
- 82: gb_hig23:*
- 83: gb_hig24:*
- 84: gb_hig25:*
- 85: gb_p11:*
- 86: gb_p12:*
- 87: gb_p13:*
- 88: gb_p14:*
- 89: gb_p15:*
- 90: gb_p16:*
- 91: gb_p17:*
- 92: gb_p18:*
- 93: gb_p19:*
- 94: gb_p1:*
- 95: gb_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1741.8	68.0	3228	9 AX050011	AX050011 Sequence
2	1659.8	64.8	2764	9 AX050010	AX050010 Sequence
3	602.4	23.5	2851	8 AF029260	AF029260 Gallus ga
4	569	22.2	3392	56 AB009864	AB009864 Expressio
5	556.2	21.7	4951	9 AR073576	AR073576 Sequence
6	554.2	21.6	184684	79 AL162732	AL162732 Homo sapi
7	552.6	21.6	215341	65 AC016904	AC016904 Homo sapi
8	534	20.9	8799	56 AF286077	AF286077 Expressio
9	478	18.7	478	9 AX050078	AX050078 Sequence
10	432.6	16.9	4614	56 AF285183	AF285183 Cloning v
11	418.8	16.4	1683	89 AK026847	AK026847 Homo sapi

Qy	774	acccggagccgcygaagagatgtaagttcttatacaacggtgcgaacgcgtggcatgtggcaa	833
Db	61	ACGGGAGGCGGAGGATGTACAGTTTTTTATCACAGGTGCAGAGCGTGGCATTTGGCAAA	120
Qy	834	gctatctgcattgtaagcagcaagaagatgtagcagaataatgtctatgtctgcgcaaaagccgc	893
Db	121	GCTATTTCATTGGAAAGCAGCAAAAGATGGAGGCMAAATTTGTTATTTGCTGCAGAAAGCCCC	180
Qy	894	cagcacacatccaaaattccacgcaacaattatactcgtctgcgcaagaatltgaagcaat	953
Db	181	CAGCACATCCAAAACTTTATAGGCACAACTATACGTGCTGGAGAAATTAAGCAGATT	240
Qy	954	ggaggaagaagccttgccaatglatgtgtgtagtgagagatggaacagagatcaagtgctca	10133
Db	241	GGAGGAAAGGCGCTTGGCCATGTATTGTTGATGTAGAGTAACACAGATCAGTCTGCA	300
Qy	1014	gtggagaagaaccatcaagaagaatttggagaagaatgtatctctgtataaatgtccagtgcc	1073
Db	301	GTGGAGAAAGCATCAAGAAATTTGGAGMAATGTATTTCTGTAAATATATGCCAGTCC	360
Qy	1074	atatagtttgaccaataatacttggaacaaacttacaagaagatggatcttgatgtaagaagt	11333
Db	361	ATTATGTTTGACCAATATACATTGGACACACCCTCACMAAGATTGGATCTGTATGTGAACATG	420
Qy	1134	aacaccagaggaacctcaactcttgatcatctaaagcaatgatatccattctgaaaaaagacaa	11933
Db	421	AACACCCAGAGCACACTCTCTTGCACTTAAAGCATGTATTTCTTATTTGAAAAAGGCANA	480
Qy	1194	gtgtctatattcccaatatacagttccaaacctgaacctgaacctaaatccagttgtgttcaaacg	12533
Db	481	GTTGCTCATATCCCAATATACGATGCCACCAGTMAACTTAAATCCAGTTTGGTTCAAAAG	540
Qy	1254	caatgtctataacacatctgctaaagtgtgatgtctatgtatgtctctggaaatggcaga	13133

QY	1374	gcctcgtgatatgcctcgtgggaagaccctggtaattcgaagaagcagctgtacgaagaagttgatatcat	1433
Db	661	GCCTTCGATATGCTGGGAGAGACCTGGTATCGAAGAGCCAGTGTGAGAAAGTTATATTCATT	720
QY	1434	gcagatgcagacataattccattttccaaaagccaaaagtttactctgccaatttgcatt	1493
Db	721	GCAGATGAGCATTATTCATTTCATTTCCAAAAGCCAAAAGTTTACTGCGCACTTTGTCAAT	780
QY	1494	gataaataatcttaaaagaagaagaaatagaataatttgaagtttatgtcaattaaacca	1553
Db	781	GATGAAATATCTTTAAAAAGAAAGGAAATGTGAAATTTTGAGCTTTATGCAATTTAAACA	840
QY	1554	ggtcatcctcttgcacaacagatttcctctttagatgtaataccagaagcagctgcagaaga	1613
Db	841	GGTCACTCTTTGCACACCAAGATTCTTCTTGTGATGAATACCAGAAAGCACTTTGCAAGAA	900
QY	1614	ctggaaatcaactctgtctgtctccagaattcaaaagaagaagaaactcgaagctgcaccca	1673
Db	901	GTCGATCACTAGTGTGCTGTCCAGAAATCAAGAAGAGAGAAACTGCAGCTGCACCAAAA	960
QY	1674	ccaagcttcctgagctctgtggaagaacatttagaattgttcaagacctcttcagtgatg	1733
Db	961	CCAGCTTCTGGAAGCTGTGGGAAGAAACATTTAGAAATGTTTAGAGACTCTCTCAGTATGAT	1020

Db 1021 GTTCTTAAAGCCACACGATCTATCTGTTTGAAGCTCTCCGGTGAAGATGGTGGCAGC 1080

Db 1081 TGGTTCTTATCTGAAAGCAGGCTGGGATGTGGATATGGAGCCTTCTGATCAG 1140

REFERENCE	LOCUS	DEFINITION	VERSION	KEYWORDS	SOURCE	ORGANISM
1259	cttgaatgagcgaagaagatttaaagggaattgcagtcacatcatatcatatgacctaaaca	1358				
716	TTGGGAATGGCAGGAAGATTTTAAAGGGAAGTGGCTGTCTAATCTTTATGCGCTTAAACA	657				
1359	gcatacacactgctgctatgatatcgtctgagagaccctgctatcgaaacagcttaga	1418				
656	GCCATTACATACAGTTCGTATGGATATCTCTGGAGAGATTGGAAATTAAGAAAGACGTCCAGA	597				
1419	aaagttgatalcatcttcagatgacgatalttccaatlttccaaagaccaaagtttact	1478				
596	AAAACTGCACATTCATGGACAGATGCTGCATATTTGACATTTTAAACAAAGCCAAAGCTTCACCT	537				
1479	ggcaacttgcctatgcttgtaaaatattctttaaagaagaagatagaaaaatttgcacgt	1538				
536	GGAACCTTCATATTATGATGAAGAGTTCTACTGAGACAGACAGGCGTTANGCATTTTNAATGTC	477				
1539	tatgcaattaaaccaaggtcatcctcttgcacacagatcttctcttagatgaatacccaaga	1598				
476	TATTCAAATTCACACAGGACA-CCCCCTGATGCCCTTACTTCTTGTGATGCTGAACCTG--	418				
1599	gcagttgcaagaagaagtgatgaacacacgtgtgtcgttccagatccaagaagaagaaactg	1658				
419	-----ACATGACAGAAAGCCATGCTGCTTCACAGGCATCAGATGAGGAGCCAAAG	373				
1659	cagctgcaccaaaacacacagcttctcgagctctggaagcaaatltagaatltgttaaagac	1718				
372	GCTATATAAGATCTGTGCTCCAGCTGGGCGCTGTTGACAGAAACATTCAGACTTATCCAGGGA	313				
1719	tctctcagttgattgttgtttaaagccactccaagcaatctatctgttgaactctccgt	1778				
312	GAACTCATATAAGAAATGGTGAAATCCACCCAGGCGCTCTTTAGTTGAACTGCTAGAGT	253				
1779	gaagatgtgtgcacgtgttcttctgactctgaagaagaagtggaatgtgcgatatgtga	1838				
252	GATGAGAGGAGGCACTTGGTACCTTGATCTGAATTAACACAGACGCGGAGCGGGG--TTTGG	195				
1839	gagcctctgtcatcagcgcagatgtgtgtatgatatgacactcatgtacattgtttaaagt	1898				
194	GAACTCTTCGACAGCCTGACGCTGTGATAGCATGTGCGAGCMGCGCTTTTGGAAGATG	135				
1899	tttcaaggaaacccaacacacacatgacatcatctgtccaaggaagaattgaagatlaaagt	1958				
134	TTCACCTGTTAACTAAAGCCACACATGCGCTTCCTCATGTGCAGAGAAATTAAGATTTAAAGT	75				
1959	aacatgtgccttagcaatcaatcattggaagact-aatgaatcaatgtaatgccagactgtg	2017				
74	AACATGCACTAGACAGTAAAGCTAGAGAAAGATGCGCTGCACACAGTTTAACCTTAACGTG	15				
2018	a					
14	A					
4	AB009864	3392 bp	DNA	circular	SYN	26-DEC-1997
AB009864	Expression vector pME185-FL3, complete sequence.					
AB009864	GI:2723416					
expression vectors plasmid: pME185-FL3 DNA.						
artificial sequence.						
1 (bases 1 to 3392)						
Maruyama, K. and Sugano, S.						
Direct Submission						
Submitted (16-DEC-1997) to the DDBJ/EMBL/Genbank databases. Sumio						
Sugano, The Institute of Medical Science, University of Tokyo,						
Department of Virology, 4-6-1, Shirokane-dai, Minatoku, Tokyo 108,						
Japan (E-mail: vsugano@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5286,						
fax: 81-3-5449-5416)						
2 (bases 1 to 3392)						
Maruyama, K. and Sugano, S.						

FEATURES	source	Location/Qualifiers	1..3392
JOURNAL	pmel18s-FL3: a versatile expression vector	Only in Database (1997) in press	
PROMOTER	polya_site	1229..1427	
BASE COUNT	810 a 902 c 818 g 862 t		
ORIGIN			
Query Match	22.2%	Score 569;	DB 56;
Best Local Similarity	96.8%	Pred. No. 3.7e-72;	Length 3392;
Matches 666; Conservative	0;	Mismatches 11;	Indels 11; Gaps 8;
1	aggcgaagtaagcaaacatgcataatcgaatagtcagcaaacatagtcgcgcgcct	60	
131	aggcgaagtaagcaaacatgcataatcgaatagtcagcaaacatagtcgcgcgcct	187	
61	aactccgccatccgcgcgccttaactccgcgcgccttaactccgcgcgccttaactccgcgcgcct	120	
188	aactccgccatccgcgcgccttaactccgcgcgccttaactccgcgcgccttaactccgcgcgcct	245	
121	tgactaatcttttattatgacga-gccgagcgcgctcgctcgactgacatccag	179	
246	tgactaatcttttattatgacga-gccgagcgcgctcgctcgactgacatccag	305	
180	aagtagtagagagagctttttgagagcctagagcttttgcaaaaagcctct-cgatcagag	238	
306	aagtagtagagagagctttttgagagcctagagcttttgcaaaaagcctctccgacgagag	365	
239	gagtcgacatctctcttaacgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgct	298	
366	gagtcgacatctctcttaacgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgct	425	
299	gagtcgacatctctcttaacgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgct	358	
426	gagtcgacatctctcttaacgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgct	485	
359	aagtttaaacgag	418	
486	aagtttaaacgag	545	
419	aactcagcgagcctcgaagccttcgacgagccttcgacgagccttcgacgagccttcgacg	478	
546	aactcagcgagcctcgaagccttcgacgagccttcgacgagccttcgacgagccttcgacg	605	
479	agtttctgcttcgagcgttacagatcccaagccttgaaaaacgaaagtaactgata	538	
606	-ctttttctgcttcgagcgttacagatcccaagccttgaaaaacgaaagtaactgata	664	
539	agtttctgcttcgagcgttacagatcccaagccttgaaaaacgaaagtaactgata	598	
665	agtttctgcttcgagcgttacagatcccaagccttgaaaaacgaaagtaactgata	721	
599	actgctcctcagtgagatgttgcttaactcctcaggagccttgagagagtgtaactctgct	658	
722	actgctcctcagtgagatgttgcttaactcctcaggagccttgagagagtgtaactctgct	781	
659	ctaaaagcttgagagatcctaactgact	686	
782	ctaaaagcttgagagatcctaactgact	809	
RESULT	5		
AR073576	AR073576	4951 bp	DNA
LOCUS	Sequence 1 from patent US 5952171.		PAT
DEFINITION	AR073576		28-AUG-2000
ACCESSION	AR073576.1	GI:10000340	
KEYWORDS			


```

misc_feature      6320..6449 /note="vacuolin thymidine kinase gene cluster 3' end"
rep_origin        6433..7047 /note="similar to pUC19"
misc_feature      7050..8192 /note="p15A intermediate copy number"
CDS               complement(7192..8052)
                  /codon_start=1
                  /product="beta lactamase"
                  /protein_id="A045964.1"
                  /db_xref="GI:12034713"
                  /translation="MSIQHRYVALIPEFAFLCPVFAHPETLVKVAEDDLGARVY
                  IELDLSGKILSFRRFRRPMSTFKVLGAVLSRIQSGDGLRIHRSQNDLVE
                  YSPVTEKHLTDGTVRELSAATITMSDNTAAILLTIGSPKELTAFIHMGGHVRLE
                  DWPEPELNEAIPNDERDITPVAMATLRLLGELLTSLRQLLIDMWADYVAGPL
                  LRSALPAGWFIADKSGAGERSGIIAALPDGKPSRIVIVYTTGSQATMDERNROIA
                  ELGASLIKHW"
misc_feature      8210..8636
BASE COUNT       2343 a 2028 c 2168 g 2260 t
ORIGIN
Query Match      20.9%: Score 534; DB 56; Length 8799;
Best Local Similarity 95.9%: Pred. No. 3.6e-67;
Matches 649; Conservative 0; Mismatches 6; Indels 22; Gaps 9;
QY 1 aggcagaagatgcaagcatgcatcctcaaatagtcagcaaacatagtcgcgcgcct 60
    |||||
DB 639 AGGCAGAGATGCAAGCATGCAATGTC-AATTAGTCAGC-AACCATAGTCC-GCCCT 695
QY 61 aactcgcgcacccgcgcctactccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 120
    |||||
DB 696 AACTCCGCCCATCCGCCCTTACTCCGCC--AGTCCGCCCATCTCCGCCCATGCG 753
QY 121 tgactaattttttattatgacaga-gccagagcgcgcgcgcgcgcgcgcgcgcgcgc 179
    |||||
DB 754 TGACTAATTTTTTATTATGACAGAGCGAGCGCGCTCGCCTGAGCTATTCCAG 813
QY 180 aagtagagagagagccttttggagcctagagcctttgcaaaaagcctc-cgactcag 238
    |||||
DB 814 AAGTAGAGAGAGGCTTTTGGAGGCTAGGCTTTGGCAAAAAGCTCTCCGATCCAG 873
QY 239 ggcctgcacatcctcctcaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 298
    |||||
DB 874 GGCCTGCATCTCTTCCAGCGCGCGCGCGCTTACTGAGCGCGCATCCAGCGCGTT 933
QY 299 gagtcgagtcctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 358
    |||||
DB 934 GAGTCGCGTTCTGCGCGCTCCCGCTGCGCTGCTCTGAACTGCTCCGCGCTAGGT 993
QY 359 aagtttaagcctcagtcgagcagcgcctttgctgcgcgcctcctctggagcctcactag 418
    |||||
DB 994 AAGTTTAAAGCTAGGTCGAGACCGGCGCTTTGTCGCGCGCTCCCTTGGAGCTTACTAG 1053
QY 419 actcaagcgcgcctcgaagccttgcctgaagccttgcctgaactctacgctcttgcctc 478
    |||||
DB 1054 ACTCAGCGCGCTTCCAGCGCTTTCCTGACCTCTGCTCAACTACGCTTTGTTTC 1113
QY 479 agtttctgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 527
    |||||
DB 1114 -GTTTCTGTTCTGCGCGCTTACGATCAAGCTCGCTCGAGGAACTGAAAACCAAAA 1172
QY 528 gttcaactgtaagctttagctcttgccttattcaagtcgcgcgcgcgcgcgcgcgcgcgc 587
    |||||
DB 1173 GTTAACGGAAGTTAGTCTTTTGTCTTTTATTATTCAGGTCGCCGATCC--GCTGTC 1229
QY 588 caaatcaagaactgcctcctcagtgagtgctgccttactcttaggcctgtaagaaatg 647
    |||||
DB 1230 CAATCAAAAGAACTGCTTCCAGTGATGTTGCTTACTTCTAGGCGCTGACGAGATG 1289
QY 648 ttacttctgtcttaaaa 664

```

```

DB 1290 TTACTTCTGCTTAATA 1306
|||||
RESULT 9
AX050078 478 bp DNA PAT 12-JAN-2001
LOCUS Sequence 91 from Patent MO0071710.
DEFINITION AX050078
ACCESSION AX050078
VERSION AX050078.1 GI:12226478
KEYWORDS
SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 478)
Deneffe, P., Rosler-Montus, M.F., Arnould-Reguligne, I., Prades, C. and
Clepet, C.
TITLE Expression products of genes involved in diseases related to
JOURNAL cholesterol metabolism
Patent: WO 0071710-A 91 30-NOV-2000;
Aventis Pharma S.A. (FR)
FEATURES
LOCATION/Qualifiers
SOURCE 1..478
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 154 a 99 c 104 g 121 t
ORIGIN
Query Match      18.7%: Score 478; DB 9; Length 478;
Best Local Similarity 100.0%: Pred. No. 3.8e-59;
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 796 cagttttatcacaggtgcaagcgcgtgcatggtgcaagctatgcatgtaagaacagaa 855
    |||||
DB 1 CAGTTTATACAGGTGCAAGCGCGTGGCATTTGGCAAAAGCTATGCAATGAACAGCAA 60
QY 856 agagtaggaacaatatgttatgtcgtgcaaaagccgcgcgcgcgcgcgcgcgcgcgcgc 915
    |||||
DB 61 AGGATGCGAGCAAAATTTGTATTCTGTCGAAGACCGCGCCACCATATCAAAACTTCTAG 120
QY 916 gcaaatctatactgcgcgtcgtgaagaattggaagatgttgaggagaaagccttgcacgt 975
    |||||
DB 121 GCACAACTTATATGCTGCTGCAAAATTTGAAGCAAGTTGAGGAAAGGCTTGCATGTA 180
QY 976 ttgttgatgtgagatggaacagcagatcagtgctgcagtgagagaaagccataagaat 1035
    |||||
DB 181 TTGTGATGTAGAGATGACAGACAGATCAGTGCATGTCAGTGGAGAAAGCCATCAAGAAAT 240
QY 1036 ttggaagaaatgatattcgtgtaataatgacagtcgcatagtttgaccaatacatgtg 1095
    |||||
DB 241 TTGGAGGAATGATATCTGCTGAATATGCACTGCACTAGTTTGACCAATACATTTGG 300
QY 1096 acacactaccagaagatgtgactctgatgatgaacgtgaacacacagagagcactacttg 1155
    |||||
DB 301 ACACACCTACCAAGATGTTGATCTGATGAAACGTGAACACACAGGCGACCTTACCTTG 360
QY 1156 catctaagcagatgcatcctattcttgaagaagacaaagttggtccatctcctaataatca 1215
    |||||
DB 361 CATTTAAGCATGTATTCCTTATTGAAAAAGACCAAGTTGCTCATATCTCTCAATATCA 420
QY 1216 gtccaccactgaacctaataacagtttgcctcaaacagcactgcttataaccattgc 1273
    |||||
DB 421 GTCACACACGAACCTTAATCACTTTGTTCAAAAGACACTGTCATATGCAATTCG 478
RESULT 10
AF285183 4614 bp DNA circular SYN 29-JAN-2001
LOCUS Cloning vector pSGHV0, complete sequence.
DEFINITION AF285183
ACCESSION AF285183.1 GI:12584846
VERSION

```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
COMMENT
FEATURES

Cloning vector pSGHVO.
Cloning vector pSGHVO
artificial sequence: vectors.
1 (bases 1 to 4614)
Leahy,D.J., Dann,C.E. III, Longo,P., Perman,B. and Ramyar,K.K.
A mammalian expression vector for expression and purification of
secreted proteins for structural studies
Protein Expr. Purif. 20 (3), 500-506 (2000)
20541642
11087690
2 (bases 1 to 4614)
Leahy,D.J.
Direct Submission
Submitted (06-JUL-2000) Biophysics, Johns Hopkins University, 725
N. Wolfe St., Baltimore, MD 21205, USA
Mammalian expression vector: directs expression of hgh fusion
proteins.
Location/Qualifiers
1. 4614
/organism="Cloning vector pSGHVO"
/db_xref="taxon:148881"
join(700..709,970..1037)
/note="Region: human growth hormone precursor"
join(700..709,970..1130,1340..1459,1552..1716,1969..2139)
/codon_start=1
/product="human growth hormone/octahistidine/etch virus
protease recognition site fusion protein precursor"
/protein_id="AAG59857.1"
/db_xref="GI:12584848"
/translation="MATSRRSLALFLCLPMLQESAPPTPLSLFPMIRAH
RLHDLFTQDEFEAYIPKQKYSFLNPOTSLCSSTPSRETQKSNLEL
RISLDLSOMLEVOFLRSVANSLVGASDSNVLDLKEEGTLMGRLEDSPR
TGQIFKQYTSKEDTSHNDALILKNYGLCYCFKMDVETFLRIYQCRSEYSGSGFS
GHHHHHHDDIPSESLYFQSGSRAPRAARQA"
join(1038..1130,1340..1459,1552..1716,1969..2277)
/product="human growth hormone/octahistidine/etch virus
protease recognition site fusion protein"
2140..2163
/note="Region: octahistidine tag"
2164..2205
/note="Region: tobacco etch virus protease recognition
site"
misc-feature
2233..2284
/note="multiple cloning site"
complement(3506..4366)
/codon_start=1
/product="beta-lactamase"
/protein_id="AAG59856.1"
/db_xref="GI:12584847"
/translation="MSIQFVALIPFPAFCPLPVFAHPETLVKVKAEDELQGARVY
IELDNSKILLESFPERFERPMSTFKVLLGAVLSRIDAGEEQGRIHQSNDLVE
YSPVKEHLDTGMVRELCSAIRTMSDNTALLLTGGPELFAFHNMGDHYRL
DRWEPLNEAIPNDERDFTMPVAMATTLRLKLLGLLASQQLIDMDEADKVGPL
LRSAIPACMFADKSGAGERSGRIGIALLGPDGKRSRIYVITYTSQATMDERNQIA
ELGASLKNH"
BASE COUNT 1099 a 1267 c 1147 g 1101 t
ORIGIN

Query Match 16.9%; Score 432.6; DB 56; Length 4614;
Best Local Similarity 97.7%; Pred. No. 1e-52;
Matches 502; Conservative 0; Mismatches 5; Indels 7; Gaps 6;

1 aggcagaagatgcaagaatgcatcaatgaatgcaagaacgaatgacggagccct 60
|||||
131 aggcagaagatgcaagaatgcatctc-aattagtcagc-aaccctatgccc-gccccc 187
|||||
61 aactccgacatccgacccctaaactccgacccgacccatctccgaccccatg 120
|||||
188 aactccgacatccgacccctaaactccgacccgacccatctccgaccccatg 245
|||||

QY 121 tgactaatctttcttattatgacaga-gccgagccgcctcgctctgactatccag 179
|||||
DB 246 tgactaatctttttttattatgacagagccgagccgcctccgacctgactatccag 305
|||||
QY 180 aagtagtgagagagcctttttgagagccagccttttgaaaaaactcctgactagag 239
|||||
DB 306 aagtagtgagagagcctttttgagagccagccttttgaaaaaactcctgactagag 365
|||||
QY 240 gctcgacatctcctctcaacgagccgcgcctctactgagagccgacatccagcagctg 299
|||||
DB 366 gctcgacatctcctctcaacgagccgcgcctctactgagagccgacatccagcagctg 425
|||||
QY 300 agtcgctctcgcgcctcgcgcctcgcgcctcgcgcctcgcgcctcgcgcctcgcgc 359
|||||
DB 426 agtcgctctcgcgcctcgcgcctcgcgcctcgcgcctcgcgcctcgcgcctcgcgc 485
|||||
QY 360 agttcaaacgctcagctcagcagcagcagccttgcgcgcgcctcgcgcgcctcgcgc 419
|||||
DB 486 agttttaaagctcagctcagcagcagcagccttgcgcgcgcctcgcgcgcctcgcgc 545
|||||

QY 420 ctcagccgagctctcagcagccttgcgcgcgcctcgcgcgcctcgcgcgcctcgcgc 479
|||||
DB 546 ctcagccgagctctcagcagccttgcgcgcgcctcgcgcgcctcgcgcgcctcgcgc 604
|||||
QY 480 gttctctgctcgcgcgcctcgcgcgcctcgcgcgcctcgcgcgcctcgcgcgcctcgc 513
|||||
DB 605 gtttctctgctcgcgcgcctcgcgcgcctcgcgcgcctcgcgcgcctcgcgcgcctcgc 638
|||||

RESULT 11

LOCUS AK026847 1683 bp mRNA PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ23194 f1s, clone REC00490.
ACCESSION AK026847
VERSION AK026847.1 GI:10439802
KEYWORDS oligo capping; f1s (full insert sequence).
SOURCE Homo sapiens primary epithelial cells of human renal proximal
tubule cDNA to mRNA, clone_11b:REC clone:REC00490.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,
Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Tanigami,A.,
Fujitara,T., Ono,T., Yamada,K., Fujii,Y., Ozaki,K., Hiroe,M.,
Ohmori,Y., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shihahara,T.,
Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEDO human cDNA sequencing project
JOURNAL unpublished (2000)
REFERENCE 2 (bases 1 to 1683)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shihahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
Sugano, Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center;
Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna@leims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)

COMMENT

NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan: cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).

FEATURES

source
Location/Qualifiers
1..1683
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="primary epithelial cells of human renal
proximal tubule"
/clone="REC00490"

OY	240	gcgcacatctctctcttaacgagccgcgcgcctacactgaagcgcacatcaacgacggttg	299
Db	359	gctgcacatctctcttcttcaacggcccgccgcgcctacactgaagcgcacatcaacgacggttg	418
OY	300	agtcgcgtctctgcgcgcctccgcgcctctgtagtgactctgaactgcgtccgcgcgtctagta	359
Db	419	agtcgcgtttctgcccgcctccgcgcctctgtagtgactctgaactgcgtccgcgcgtctagta	478
OY	360	agtttaaaagcctcagtgtagacagcggagcctctgtccgcgcgcctcttgtagagctaccta	419
Db	479	agtttaaaagcctcagtgtagacagcggagcctctgtccgcgcgcctcttgtagagctaccta	538
OY	420	ctcagcgcgctctccaaagccttgccttgacccctgctctcaactcaagtcctttgtttca	479
Db	539	ctcagcgcgctctccaaagccttgccttgacccctgctctcaactcaagtcctttgttttc	597
OY	480	gttttcgcttctgcgcgcttacaatc	506
Db	598	gtttttctgtctgcgcgccttacaatc	624

RESULT	14				
LOCUS	AX070432				
DEFINITION	AX070432				
ACCESSION	AX070432				
VERSION	AX070432.1				
KEYWORDS	GI:12580217				
SOURCE					
ORGANISM	human.				
REFERENCE	human.				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 390)				
TITLE	Williams, L.T., Escobedo, J., Innis, M.A., Garcia, P.D., Klinger, J., Kassam, A., Reinhard, C., Randazzo, F., Kennedy, G.C., Pot, D., Lamson, G., Drmanac, R., Cirenjakov, R., Drmanac, S., Dickson, M., Labat, I., Leshkowitz, D., Kita, D., Garcia, V. and Strache-Crahn, B				
JOURNAL	Human genes and gene expression products				
FEATURES	Patent: WO 0102568-A 904 11-JAN-2001;				
CHIRON CORPORATION (US); HISEQ, INC. (US)					
Location/Qualifiers					
1..390					
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
129 a	68 c	81 g	112 t		
BASE COUNT					
ORIGIN					

	Query Match	Best Local Match	Similarity	99.5%	Pred. No.	2.Be-44	DB	10	Length	390
	Matches	375	Conservative	0	Mismatches	2	Indels	0	Gaps	
QY	1226	gaacctaaatccagcttctggtccaacagcactgctcatatccacttgcataagtagt	1285							
Db	14	GAGCCTAAATCCAGTTTGCTTCAAAACACTACTGGTCTTATCCACTTGGTAAGTGGAT	73							
QY	1286	gtctatgtagtgccttggaaatgagcagaagaatttaaagtgaaattgcaatgcaatg	1345							
Db	74	GCTATGATGATGCTGCTTGAATGCGAGAGAAATTTAAAGTGAATGCAATGCAATGCAAT	133							
QY	1346	atggcctaaaacacgcacatacacactgctgcctatgtagatagctctggagagactgtag	1405							
Db	134	ATGGCTTAAACAGCGCAATACACCTGCTGCTATGGATATGCTGGAGAGACCTGGTATCGA	193							
QY	1406	aagccagtgtagaaagtgtgataatcatctggagatgagcagcatcttccactttccaaaagcc	1465							
Db	194	AAGCCAGTGTGAGAAAGTTGATATCATTTGCGAGATGCGACCAATATTCATTTCCAAAAGCC	253							
QY	1466	aaaagatttcaatcggcaactctgtcatctgtagtgaataatctttaaagaagaagaataga	1525							
Db	254	AAAAAGTTTACTGGCAACTTTGTTCATTTGATGAAGAAATATCTTTAAAGAAGAAGGAATGA	313							
QY	1526	aaatttgaagcttatgcaattaaacagcgatcctcttgcacaacagaattctctctaga	1585							

Db	314	AAATTTTACGTTTATGCAATTAACCGATCCTTTCGACACAGTTTCTTCTTAGA	373
Qy	1586	lgaatccccagaagcag	1602
Db	374	TGAATACCCAGACGAG	390

RESULT	15
AK000572	
LOCUS	AK000572 1755 bp mRNA PRI 22-FEB-2000
DEFINITION	Homo sapiens CDNA FLJ20565 fis, clone REC00542.
ACCESSION	AK000572
VERSION	AK000572.1 GI:7020758
KEYWORDS	oligo capping; fis (full insert sequence).
SOURCE	Homo sapiens primary epithelial cells of human renal proximal tubule CDNA to mRNA, clone_1lb:RBC clone:REC00542.
ORGANISM	Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (sites)
Taniguchi, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K.,
Hirao, M., Ohmori, Y., Ota, T., Suzuki, Y., Obayashi, M., Nishii, T.,
Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T., and Sugano, S.
NEDO human cDNA sequencing project
unpublished (2000)
2 (bases 1 to 1755)
Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishii, T., Isogai, T.,
Shibahara, T., Tanaka, T., and Nakamura, Y.
Direct Submission
Submitted (15-FEB-2000) to the DDBJ/EMBL/GenBank databases. Sumio
Sugano, Institute of Medical Science, University of Tokyo, Department
of Virology, Shirokane-dai, 4-6-1, Minato-Ku, Tokyo 108-8639, Japan
(E-mail: cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan: cDNA full insert
sequencing: Research Association for Biotechnology: cDNA library
constructing, 5' - 3' and one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).

FEATURES	SOURCE	Location/Qualifiers
		1..1755
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/cell_type="primary epithelial cells of human renal proximal tubule"
		/clone="REC00542"
		/clone_id="REC"
		/note="cloning vector pME18SFL3"
		503..1489
		/note="unnamed protein product"
		/codon_start=1
		/protein_id="BAA91263.1"
		/db_xref="GI:7020759"
		/translation="MSMLERRRKKOKNAVDPONTAMNSDSKFGORMLEKMGWSKGGK LGAEOGACDPIKRYOVKRNHILGATINENNNYLAHODPEMLLAELNTGCOETPDS SDKEKKSFSLEEKSKISKRNRYHMKFTFKGDLSRSKTDIDCLFGKROSKRTPEGDD SPTEPENETTTTSAFTIOEYFAKRMALNNKPPVPEGSDISETOVERKRSKKRK ATGVDSEYTLDPKAKRHTEGPEPERAEOERVAKRSKSAPEEQILGCPGDSSKASAO AGDHVOPEEGHDTFLKPKRRGKKLKPVEIAADATLEETLVKKKKKKDSK"
BASE COUNT	490 a	449 c 464 g 352 t
ORIGIN		
Query Match	13.7%	Score 351.4; DB 88; Length 1755;
Best Local Similarity	98.1%	Pred. No. 4.1e-41;
Matches 366; Conservative 0; Mismatches 6; Indels 1; Gaps 1.		


```
QY 374 gtcgagaccgggacctgtgtccggcgctcccttgagagcttaactagactcaagccgctctc 433
    |||
DB 61 gtcgagaccgggacctgtgtccggcgctcccttgagagcttaactagactcaagccgctctc 120
    |||
QY 434 caagcttgccttgaacctgtgtcaacttaagcttgccttgccttgccttgccttgccttgc 493
    |||
DB 121 CACGCTTTGCTGACCTGCTGCTCACTCTAGTCTTGTTC-GTTTCTGTCTGC 179
    |||
QY 494 gccgttacagatccaaagctctgaaaaaccagaaagtaactgylaagttagtccttttg 553
    |||
DB 180 GCCGTTACAGATCCAAAGCTCTGAATAACCAAGTTAACTGGTAAGTTAGCTTTTG 239
    |||
QY 554 tccttaattcaagtcgccggaatccggtgtgtgcaaatcaagaactgctcctcagttg 613
    |||
DB 240 TCTTTTATTTCAAGTCCCGATCCGGTGTGTGCAAAATCAAGAACTGCTCAGTGG 299
    |||
QY 614 atgtgccttacttctagagcctgtaaggaagtgtaacttctgctctaaagctgcgaa 673
    |||
DB 300 ATGTTGCTTTACTTCTAGGCTGTACGGAAGTGTACTTGTCTTAATAAGCTGCCGAA 359
    |||
QY 674 ttctaatacgaact 686
    |||
DB 360 TTCCTCGAGCACT 372
```

Search completed: June 12, 2001, 11:40:15
Job time: 7978 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 12, 2001, 11:52:03 ; Search time 3231.71 Seconds

(without alignments)
11687.084 Million cell updates/sec

Title: US-09-464-039-8

Perfect score: 2561
Sequence: 1 aggcagagaagtatgcagaagca.....gsgmgtrasmwawwrammc 2561

Scoring table: OLIGO_NUC
Gapop 60.0 , Capext 60.0

Searched: 1283235 seqs, 7373929652 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	1305	51.0	3228	9 AX050011	AX050011 Sequence
2	1223	47.8	2764	9 AX050010	AX050010 Sequence
3	478	18.7	478	9 AX050078	AX050078 Sequence
4	418	16.3	184684	79 AL162732	AL162732 Homo sapi
5	418	16.3	215541	65 AC016904	AC016904 Homo sapi
6	348	13.6	390	10 AX070432	AX070432 Sequence
7	332	13.0	4614	56 AF285183	AF285183 Cloning v
8	305	11.9	1683	89 AK026847	AK026847 Homo sapi
9	249	9.7	3392	56 AB009864	AB009864 Expressio
10	249	9.7	4951	9 AR073576	AR073576 Sequence
11	249	9.7	8799	56 AF286077	AF286077 Expressio

	Db	1141	GCAGATGCGTGATGACGATGACTCTATCATGCTTTGTAAAAATGTTTTTCAGGGAAACTA	1200
Qy	1914	aaaccaacatgcatcattcgtlcagggaattgaagattaagaagtacaatggcccttagca	1973	
Dd	1201	AAACCACAAGGCGCTTCAATGTCAGGGAAAATTGAAGATTAAAGTAACATGCGCCCTAGA	1260	
Qy	1974	atcaaatgaggaagactaatcaatgaagttgaatgcgaaccttga	2018	
Dd	1261	ATCAAATTGGAGAAAGCTAATGATCAGATGATGCCAGACTGTGA	1305	
RESULT	2			
AX050010		AX050010	2764 bp	DNA
LOCUS		Sequence 23 from Patent WO0071710.	PAT	12-JAN-2001
DEFINITION		AX050010		
ACCESSION		AX050010.1	GI:12226383	
VERSION				
KEYWORDS				
SOURCE		human.		
ORGANISM		Homo sapiens		
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE		Denefle,P., Rosier-Montus,M.F., Arnould-Reguinne,I., Prades,C. and Clepet,C.		
JOURNAL		Expression products of genes involved in diseases related to cholesterol metabolism		
FEATURES		Patent: WO 0071710-A 23 30-NOV-2000;		
source		Avantis Pharma S.A. (FR)		
BASE COUNT		Location/Qualifiers		
ORIGIN		1..2764		
		/organism="Homo sapiens"		
		/db_xref="taxon:9606"		
		917 a 479 c 565 g 803 t		
Query Match		Best Local Similarity	47.8%; Score 1223; DB 9; Length 2764;	
Matches 1223; Conservative		0; Mismatches	0; Indels	0; Gaps
Qy	796	caattttatcacagtcgcaagcgctgcatgttggaagaacttatgatggaagaccaa	855	
Dd	1	CAGTTTATATCACAGGTGCAGCCCTGGCATTTGGCAAAAGCTATTCATTTGAAGACGCA	60	
Qy	856	aggatlgagacaatatctgttatctgtcgaagaaccgccagccacatccaacttctag	915	
Dd	61	AGGATGGAGCAAAATATTTGTTATTCCTGCAAAAGACCCGCCACATCCAAAACTTCTAG	120	
Qy	916	gacaaatctaactcgtcgtcgtcgtgaagaattgaagcagttggagagaagccttgcaatga	975	
Dd	121	GCACATCTATACTGCTCTCTGAAAGAAATTTGAAGCAGTTGGAGGAAAGGCTTGCATGTA	180	
Qy	976	tctgtgatgtgagagatgaacagcagatcagctgctgcagttggaagaagccatcaagaat	1035	
Dd	181	TTGTGTGATGTGAGATGATACACGACGATCAGTGTCTGCACTGGAGAAAGCCATCAAGAAAT	240	
Qy	1036	tltgagaagaattgatalctctgtgtaaalaylccycagtygcaattagttgaccaatagtg	1095	
Dd	241	TTGGAGCAATTGATATTCCTGTGTAATAATATGACAGTCCATFATGTTGACCAATACATTGG	300	
Qy	1096	aacaaccctcaagaagatcttgatctgcatgataagctgaacccaagagccactactcttg	1155	
Dd	301	ACACACCTAACCAAGAGATTGATCTGATGATGAACGTGAACACCAAGAGGCACCTTACCTTG	360	
Qy	1156	cattcaagaatgtatcttccatttggaaaagaagcaaeqtlbtcatatcccaatlatca	1215	
Dd	361	CATCTAAAGCATGATTCCTTATTTGAAAAAAGCAAGTAGTTCCTCATATCCICCAATATTA	420	
Qy	1216	gtccaacactgaacctaaataaccagtttgttcaaacagcactgtytcttataccattgcta	1275	
Dd	421	GTCACCACTGAACCTAATTCAGTTTGGTTTCAAAACAGACAGCTGCTTATTCACATTGGTA	480	

OY	1276	agcctgctatgctctatgctatgctgcttggaatgagcaagaatttaaagtgaaattgacg	1335
Db	481	AGTATGGATATGCTCTATGTATGTGCTTGGAAATGGCAAAATTTTAAAGGTGAATTGGAC	540
OY	1386	tcaatgcatatgagcctaaacagccatacacactgctctatgatatgctctggaagac	1395
Db	541	TCAATGCATATTAGGCTTAATAACAGCCATACACACTCTCTATGGAATTCCTGGAGAC	600
OY	1396	ctggtatcgaaagccagtgtagaaaagtgtatatacatgtcagaatgcagaatltcatt	1455
Db	601	CTGGTATCGAAAGCCAGTGTAGAAAAGTTGATATCATTCAGATGCAGCATATTCCATT	660
OY	1456	tccaaagcccaaaagtttactgacacttgtcatatgataaaatcttaagaag	1515
Db	661	TCCAAAAGCCAAAAGTTTACTGTGCAACTTGTGCTATGATAAAATCTTAAAGGAAG	720
OY	1516	aaggaatagaanaatttgcagcttatagcattaaaccaggtcatccttgcaccagatt	1575
Db	721	AAGGATAGAAAATTTTGTGCGTTTATGCATTTAAACCAGTCACTCTTCCAAACGAT	780
OY	1576	tctcttagatgaataaccagaagcagtttagcaagaagaatgynaatcaactggtgttc	1635
Db	781	TCTTCTTGAATGAATTAACCCAGAAAGCGATTAGCAAGAAATGGAATCAACTGCTGCTTC	840
OY	1636	cagaattccaagaagaagaacatgcagctgtcacaaccaaacacagttcttgagctgtggaag	1695
Db	841	CAGAAATTCAAAGAAAGAAATGCACCTGCACCAACCAACCTTGTGAGCTGTGGAG	900
OY	1696	aaacattagaattggtlaagagactcctcagtgatgattgtgtlaaagccactcaagca	1755
Db	901	AAACATTTTGAATTTGTTAAGCATCTCTCAGTATGATGTTGTTTAAAGCCACTCAAGCA	960
OY	1756	tctatctgttgaactcctccggtgaagaatggtgagcagtggtttcttgcattcgtaaaaagca	1815
Db	961	TCTATCTGTTTAAACCTCCCGGTGAAGATGTGGCAAGTGTTCTTGATCTGAAAAACA	1020
OY	1816	agggtgggaatgtcgatatatgagagaccttcctgataagcagatgtygtgtagatga	1875
Db	1021	AGGGTGGGAATGTCCGATATTGGAGAGCCTTCTGATATGAGCAGATGTGGTATAGATATGA	1080
OY	1876	ctactgatatcttgttgaataatggttttcagggaaactaaaccaaatgycatltcatt	1935
Db	1081	CTACATGATACATTTGTAAAAATGTTTTCAGAGCAAACTTAAACCAACAAATGCAATTCATGT	1140
OY	1936	cagggaatattgaagattlaaagttaaatgtgccttagcaatcaaatgtggagaactatga	1995
Db	1141	CAGGGAATATTGAAGATTAAAGGTAAACATGTGCTTACCAATCAAAATTTGGAGAACCTAATGA	1200
OY	1996	atcagatgaatgccagactgtga	2018
Db	1201	ATCAGATGAAATGCCACACTGTGA	1223
RESULT	3		
AX050078			
LOCUS	AX050078	478 bp	DNA
DEFINITION	Sequence 91 from Patent WO0071710.		PAT
ACCESSION	AX050078		12-JAN-2001
VERSION	AX050078.1	GI:12226478	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1 (bases 1 to 478)		
	Deleffe,P., Rosier-Montus,M.F., Arnould-Reguigne,I., Prades,C. and		
	Clepet,C.		
TITLE	Expression products of genes involved in diseases related to		
	cholesterol metabolism		
JOURNAL	Patent: WO 0071710-A 91 30-NOV-2000;		
FEATURES	Location/Qualifiers		

```
source 1.478
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 154 a 99 c 104 g 121 t
ORIGIN

Query Match 18.7%; Score 478; DB 9; Length 478;
Best Local Similarity 100.0%; Pred. No. 1.8e-252;
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 796 cagtttttaccacagtgacgacgctgacgttgacaaagctatgcatgaaagcagca 855
|||||
DB 1 CAGTTTATACAGAGTGCAAGCGCTGGCATTTGGCAAAAGCTATTGCAATGCAAGCA 60
|||||
QY 856 aggatgagagaatattgtattgtctgacaaagccgacgacacatcaaaacttag 915
|||||
DB 61 AGGATGAGCAAAATTGTTATTGCTGCAAAAGCCGCCACATCCAAACTTCTAG 120
|||||
QY 916 gcaaatctatctgctgctgaaagaaatgaaagcttgaggaaaggccttgcatgta 975
|||||
DB 121 GCACAAATCTATTACTGCTGCGAAGAAATGAAACAGCTTGAGGAAAGCCTTGCAATGTA 180
|||||
QY 976 ttgttgatgagagatgaaacagacatgctgctgacgtggagaaagccatcaagaat 1035
|||||
DB 181 TTGTTGATGAGAGATGAACACAGATCAGTCTGCAGTGAGAAAGCCATCAAGAAAT 240
|||||
QY 1036 ttgaggaatctatctgtaataatccagtgacatgattgacacatcatctg 1095
|||||
DB 241 TTGAGAGAAATGATATTCTGCTGAATAATGCCAGTGCCATTAGTTACAAATACATTTGG 300
|||||
QY 1096 acacacacacacagagatgagctgcatgtaagcgtgaaacacacagagacacacac 1155
|||||
DB 301 ACACACCTACCAAGAGATGTGATCTGATGTAACGGAACACACAGACGACCTACCTTG 360
|||||
QY 1156 catctaaagatctatctcatttgaaagaaaggaagtgcctcatctcaatctca 1215
|||||
DB 361 CATCTAAGAGATGATCTCTTATTGAAAAGCAAGATTGCTCATATCTCAATATCA 420
|||||
QY 1216 gtccacacacacacacacacacacacacacacacacacacacacacacacacac 1273
|||||
DB 421 GTCCACACACACACACACACACACACACACACACACACACACACACACACACAC 478
|||||

RESULT 4
AL162732 184684 bp DNA HTG 07-FEB-2001
LOCUS Homo sapiens chromosome 9 clone RP11-32M23, *** SEQUENCING IN
DEFINITION PROGRESS ***, 3 unordered pieces.
ACCESSION AL162732.28 GI:12717974
VERSION AL162732
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 184684)
REFERENCE 1
AUTHORS Sehra,H.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Feb 8, 2001 this sequence version replaced g1:12597074.
COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA32M23
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M7815; 0% of reads
```

```
Sequencing vector: plasmid; L08752; 99% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 18441 bases at least Q40
Consensus quality: 18443 bases at least Q30
Consensus quality: 18463 bases at least Q20
Insert size: 18464; sum-of-contigs
Insert size: 186775; 9.9% error; agarose-fp
Quality coverage: 13.53x in Q20 bases; sum-of-contigs Quality
coverage: 13.46x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 45498: contig of 45498 bp in length
* 45499 45598: gap of 100 bp
* 45599 108237: contig of 62639 bp in length
* 108238 108337: gap of 100 bp
* 108338 184684: contig of 76347 bp in length.
*
FEATURES
source
1.184684
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-32M23"
/clone_1lb="RPCT-11.1"
1.45498
/note="assembly-fragment:06222
fragment_chain:1"
45599.108237
/note="assembly-fragment:02558
fragment_chain:1"
108338.184684
/note="assembly-fragment:00830
fragment_chain:1"

BASE COUNT 52625 a 36409 c 38646 g 56804 t 200 others
ORIGIN

Query Match 16.3%; Score 418; DB 79; Length 184684;
Best Local Similarity 100.0%; Pred. No. 5.5e-219;
Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2045 aggaataataaaaaaagacgacgctgctgctcaaaaagtaaaaaagctcaagct 2104
|||||
DB 91849 AGGAATAATAAAAAAAGTCACTGCTATGCTCAAAAAGTA AAAAGCTCAACAGTT 91908
|||||
QY 2105 aaaaatcaatgctgtcttcttctcgttataataaagataatgacgtttctctgga 2164
|||||
DB 91909 AAATCTAATGTTGTTGTTCTTCCTGTTATTAATGAAGATATGACGTTGTTGGA 91968
|||||
QY 2165 aaagatagaattgctctcaaaagacttgaaattgtaataaaatggaagcaatcaaa 2224
|||||
DB 91969 AAAGATGAATTTGCTCTAAAGACTTGAATTAATGAATGAAGCAAGCTAATCAAA 92028
|||||
QY 2225 cataagctcattaaagtggagacttaagacagctcgttcttataatccaaggttaa 2284
|||||
DB 92029 CATTAAGCTTATTAAAGTGGATTTCTAAGACAGTCTGTGTTTATTATTTCAAGGTTTAA 92088
|||||
QY 2285 cccttgagccttacctatctatctactgctctctccaagaagaatatttggcgaca 2344
|||||
DB 92089 CCCTTGAGCGCTTAACATCTCATCTGCTCTCTCCAGAAAAGTATTGGCGGACAC 92148
|||||
QY 2345 gtcaagatcaagcagtaaatatgactcttccaatctcttgtaatgtaaaatgaagctag 2404
|||||
DB 92149 GTCAAGATCAAGCAGTAATAATTAAGCTCTTTCAAATCTTCTGATGATGAATAAAGCTAG 92208
|||||
QY 2405 tctgtttaaatttttagtttggttggtgtaataaagtaaaatcttaagaagcttt 2462
|||||
```


Db	4969	CCCTTGGAGCCTTACATGTCATTCACTGTCCTTTCCAAAGAANAAGTAATTTGGGGGACA	4910
Oy	2345	gtccagatcgaagcgataaaattatgccttccttcaaatctttctgcatgtaaatgaagctcag	2404
Db	4909	GTCACATCATGAAGAGTAATAATTAGCTCTTTCAAATCTTCTGTGCATGTAAAAAAGAGCTAG	4850
Oy	2405	tctgtcttaaaacttttagtttgattgatgtatcaactgaataatctaagtcttc	2462
Db	4849	TCTGTTTTAAATTTTGTAGTTTGATTTGGATTGTATACTAAGAANAATCTTAATGATGTTTT	4792
RESULT	6		
AX070432			
LOCUS	AX070432	390 bp	DNA
DEFINITION	Sequence 904 from Patent WO0102568.		PAT
ACCESSION	AX070432		
VERSION	AX070432.1	GI:12580217	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
	1 (bases 1 to 390)		
	Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D., Klinger,J.,		
	Kassam,A., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,		
	Lasson,G., Dmanac,R., Girenjakov,R., Dmanac,S., Dickson,M.,		
	Labat,I., Leshkowitz,D., Kita,D., Garcia,Y. and Strache-Crain,B.		
	Human genes and gene expression products		
	Patent: WO 0102568-A 904 11-JAN-2001.		
	CHIRON CORPORATION (US); HYSIQ, INC. (US)		
FEATURES			
Source	Location/Qualifiers		
	1..390		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	129 a	68 c	81 g
ORIGIN			112 t
Query Match	13.6%	Score 348;	DB 10; Length 390;
Best Local Similarity	100.0%;	Pred. No. 1,4e-180;	
Matches 348:	Conservative	0; Mismatches	0; Indels
			0; Gaps
Oy	1255	actggtcctatccattcctaagaatgtagtgcattcatgtatgtcttggaatgagcagaag	1314
Db	43	ACTGTCCTTATTCACCTTGTCTAGTAGTATGATGTCTATGATGTCTTGGAATGCAGAAAG	102
Oy	1315	aattcaagglyaaattcgcagtcacatgcatatgagcctaaacagccatacacactgcty	1374
Db	103	AATTTAAAGGTGTAATTGCGAGTCAATGCTATTGTGGCTTTAAACACCACATACACACTGCTG	162
Oy	1375	cctagtatatgtctggagaagacctggtatcatgaaagccagctgtagaagaagttagatcatg	1434
Db	163	CATATGATATGCTGGGAGAGCACTGATTCAGAAAGCCAGTGTAGAAAAGTTGATATCATTTG	222
Oy	1435	cagatgagcagcatatccatttcaccaaaagccaaaagtttcaactltygcaactltgcatgy	1494
Db	223	CAGATGACAGCAATATTCATTTCACAAAAGGCCAAAAGTTTACTGCAACTTTGTCATTG	282
Oy	1495	atgaaaaatactttaaagaagaagaatagaanaatttgagcgttatgtacaattaaccag	1554
Db	283	ATGAAATAATCTTTAAAGAAAGAAAGAAATGAATAATTTTGACCTTTATGCAATTAACACAG	342
Oy	1555	gtcatccttgcacacagatctctcttgaatgaatlacccagaagcag	1602
Db	343	GTCATCCTTTGCCAACCAAGTTTCTTCTTAGATGAATACCCAGAACAG	390
RESULT	7		
AF285183		4614 bp	DNA
LOCUS	AF285183		circular SYN
DEFINITION	Cloning vector pSGH0, complete sequence.		
ACCESSION	AF285183		
VERSION	AF285183.1	GI:12584846	
			29-JAN-2001

KEYWORDS	Cloning vector pSGHV0.
SOURCE	Cloning vector pSGHV0.
ORGANISM	artificial sequence; vectors.
REFERENCE	1 (bases 1 to 4614)
AUTHORS	Leahy,D.J., Dann,C.E. III, Longo,P., Perman,B. and Ramyar,K.X.
TITLE	A mammalian expression vector for expression and purification of secreted proteins for structural studies
JOURNAL	Protein Expr. Purif. 20 (3), 500-506 (2000)
MEDLINE	20541642
PUBLISHED	11087690
REFERENCE	2 (bases 1 to 4614)
AUTHORS	Leahy,D.J.
TITLE	Direct Submission
JOURNAL	Submitted (06-JUL-2000) Biophysics, Johns Hopkins University, 725 N. Wolfe St., Baltimore, MD 21205, USA
COMMENT	mammalian expression vector; directs expression of hgh fusion proteins.
FEATURES	Location/Qualifiers
source	1..4614
orgnism	"Cloning vector pSGHV0"
db_xref	"taxon:148881"
join	(700..709,970..1037)
join	(700..709,970..1130,1340..1459,1552..1716,1969..2139)
note	"Region: human growth hormone precursor"
join	(700..709,970..1130,1340..1459,1552..1716,1969..2280)
codon_start	1
product	"human growth hormone/octahistidine/etch virus protease recognition site fusion protein precursor"
protein_id	"AAG59857.1"
db_xref	"GI:12584848"
translation	"MATGSKTSLLLAFGLCLPWLQESAPFTPLPLSLFEDNAMIARHRLQALADYQEEFEAYIPKEQKYSFLDNQPSLCSSESIPPSNKEETQOKSNELRLISLLIQSWLEPVQFLRSYFANSLVYGASDSNVYDLDELDEGIOTLNGRSDSPRTGQLEKQYKSKPTGSHNDALKNYGLGLCEFKDKDKYETFLRIYQCSVSGSGFSGHHNNHNDYDIIPSSNELYFQSSNARPAARGHA"
join	(1038..1130,1340..1459,1552..1716,1969..2277)
product	"human growth hormone/octahistidine/etch virus protease recognition site fusion protein"
join	(140..2163)
note	"Region: octahistidine tag"
join	(2164..2205)
note	"Region: tobacco etch virus protease recognition site"
join	(2233..2284)
note	"multiple cloning site"
complement	(3506..4366)
codon_start	1
product	"beta-lactamase"
protein_id	"AAG59856.1"
db_xref	"GI:12584847"
translation	"MSIQHRVALLIPFAFCIPVAPETLVKYNDAEDQAGAVGYIELDNSGKILSEFRPEERFPMSTKVLICGAVVSRIDAGQDQGRRIHNSONDLVEYSPVTEKHNLDTGNTVRELCSAALTMSDNTNANLLITGSPLELAFINMWDHYTRLDRNEPDLMEIIPNDEKRTTTPVAMATTLLRLGLGELLTASROQLDWMEDAKVAGPLLRSLRLPGWFIADKSGAGERSGIIAALGPDKPSRIIVYITTSQATMDERNROIAEIGSLIKHM"
BASE COUNT	1099 a 1267 c 1147 g 1101 t
ORIGIN	
Query Match	13.0%; Score 332; DB 56; Length 4614;
Best Local Similarity	100.0%; Pred. No. 1.3e-171;
Matches 332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
0Y	147 gccagagcgcgcctgcgcctcctagctattccagaagatgtagagagccttttcgagac 206
DB	273 gccagagcgcgcctgcgcctcctagctattccagaagatgtagagagccttttcgagac 332
0Y	207 ctaagccttttcaaaaagctcctcgtatcgcagagagcctgcacctccttaagcgccgc 286
DB	333 cttagccttttcaaaaagcctcctcgtatcgcagagagcctgcacctccttaagcgccgc 392

0y 267 cgcctcactgagggccgcacatccacgacggttgatcgcttcgacgctcccgccgtgt 326
|||||
Db 393 CGCCTTACTGAGGCGCGCATCCAGCGGTTGAGTGCCTTCCGCTCCGCGCTGT 452
0y 327 ggtgctcctgaactgctgcgcgcgtctaggtaagttaaagctcagtcagacggcgc 386
|||||
Db 453 GGTGCTCTCAGTAACTGCGTCCGCGGTCTAGTAAAGTTAAAGCTCAGTCAAGACCGGCG 512
0y 387 ctgtgcgagcgctcccttgtagcctactagactcagcgctctccagcttgctg 446
513 CTTTGTCTCCGGGCTCCCTTGAGGCTTACTGACTCAGCCGCTCCACGCTTGTGCTG 572
Db 447 accctgcttgctcaactcactgcttgcttc 478
|||||
Db 573 ACCCTGCTTCTCACTTACTGCTTGTGTTTC 604

RESULT 8
AK026847 1683 bp mRNA PRI 29-SEP-2000
LOCUS AK026847
DEFINITION Homo sapiens cDNA: FLJ23194 fis, clone REC00490.
ACCESSION AK026847
VERSION AK026847.1 GI:10439802
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens primary epithelial cells of human renal proximal
tubule cDNA to mRNA, clone_11b:REC clone:REC00490.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (sites)
Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,
Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Tanigami,A.,
Fujiwara,T., Ono,T., Yamada,K., Fujii,Y., Ozaki,K., Hiroo,M.,
Ohmori,Y., Ota,T., Suzuki,Y., Odayashi,M., Nishi,T., Shibahara,T.,
Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
Unpublished (2000)
2 (bases 1 to 1683)
Sugano,S., Suzuki,Y., Ota,T., Odayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
Submitted (29-AUG-2000) to the DDBJ/EMBL/Genbank databases. Sumio
Sugano, Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center;
Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan: cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'-8' and one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).

FEATURES
source
1..1683
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="primary epithelial cells of human renal
proximal tubule"
/clone="REC00490"
/clone_11b="REC"
/note="cloning vector PME18SFL3"

BASE COUNT 506 a 333 c 303 g 541 t
ORIGIN

Query Match 11.9%; Score 305; DB 89; Length 1683;
Best Local Similarity 99.8%; Pred. No. 1e-156;
Matches 425; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

0y 251 tctctcagcgccgcgacctactgagcgccatccacgacggttgatcgcttcct 310
|||||

Db 7 TCCTTACAGCGCGCGCGCCCTTACTGAGGCGGCATCCAGCGCGTTCGAGTTCCT 66
0y 311 ggcgcctccgcgcgtgtgtgctcctcctgaactgctgcgcgtctcagtgtaagttaaact 370
|||||
Db 67 GCCGCTTCCCGCTCTGTGCTGCTCTGAACTGCGTCCCGCTTGAAGTTAAAGCT 126
0y 371 caggtcgagacggcgctcttcgagcgctcctccttgagacctactaactcagcgcgct 430
127 CAGGTGAGACCGGCGCTTGTGCGGCTCTCTTGAGGCTTACTGACTTAACTCAGCGGCT 186
0y 431 ctcaacgctttgctcagacctgctgtcctcaactcactgctcttgcttgcttcgctc 490
187 CTCACGCTTTGGCTGACCGCTGCTGCTCACTGCTTGTGTTTC-GTTTTCGTGTC 245
Db 491 tgcgcgcttacagatccaaactcgtgaaacacagaagttactgtaagttagcttc 550
246 TCGCGCGTTTACAGATCCCAAGCTCTGAAAAACAGAAATTAAGTTAAGTTAGTCTTT 305
0y 551 ttgtctttatttaagagcgccgagatccggtgtgtgtgtgtaaatcaagaactgctcag 610
Db 306 TTGTCTTTTATTTTTCAGGTCCGCGATCCGCTGTGTGCAAAATCAAGACTGCTCTAG 365
0y 611 tggatgtgcttactcctcctcctcagcctgacggaagtgtactctgctcctaaagctgcg 670
Db 366 TGGATGTGCTTACTTACTTACTGAGCGCTTACGGAAGTGTACTTCTGCTTAAAGCTGCG 425
0y 671 gaattc 676
|||||
Db 426 GAATTC 431

RESULT 9
AB009864 3392 bp DNA circular SYN
LOCUS AB009864
DEFINITION Expression vector PME18S-FL3, complete sequence
ACCESSION AB009864
VERSION AB009864.1 GI:2723416
KEYWORDS expression vectors plasmid:PME18S-FL3 DNA.
SOURCE Plasmid vectors
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 3392)
Matsuyama,K. and Sugano,S.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1997) to the DDBJ/EMBL/Genbank databases. Sumio
Sugano, The Institute of Medical Science, University of Tokyo,
Department of Virology; 4-6-1, Shirokane-dai, Minato-ku, Tokyo 108,
Japan (E-mail:sugano@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
2 (bases 1 to 3392)
Matsuyama,K. and Sugano,S.
TITLE PME18S-FL3: a versatile expression vector
JOURNAL Published Only in Database (1997) In press
FEATURES
source
1..3392
/organism="vectors"
/plasmid="PME18S-FL3"
/db_xref="taxon:29278"
promoter
1..638
polya_site 1229..1427
BASE COUNT 810 a 902 c 818 g 862 t
ORIGIN

Query Match 9.7%; Score 249; DB 56; Length 3392;
Best Local Similarity 100.0%; Pred. No. 1e-125;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0y 230 cgatcagagggtcgtcgtcctcctcctcagcgcgcgcctcctcctcagcgcacatcc 289
Db 357 CGATCGAGGGGCTCGCATCTCTTTCACGCGCGCGCGCGCTTACTGAGCGCGCATCC 416
0y 290 acgcggttgagtgctgcgttcgcgcgcctccgcgctgtgtgctcctcctgaactgctcgc 349

```

Db      417  ACCCGCGTGGAGCGCGTCTCGCCGCTCCGCGCTCGGCGCTCTGAGACTCGTCGCC 476
          |||
Qy      350  cgtctagtaagttaaaagctcagatcagacgagcgcccttgcgcgcgccttcggag 409
          |||
Db      477  CGCTGAGTAAGTTTAAAGCTCAGGTCGAGACCGGGCCCTTTGTCGGCGCTCCCTTGAG 536
          |||
Qy      410  cctactagactcagcgcgcctcctcagcttgcctcagacccttgcctcgaactcga 469
          |||
Db      537  CCACTAGACTCAGCGCGCTCTCAGAGCTTTCCTGACCTGCTGCTCACTACGTA 596
          |||
Qy      470  cttgttctc 478
          |||
Db      597  CTTTGTTTC 605

RESULT 10
AR073576 LOCUS AR073576 4951 bp DNA PAT 28-AUG-2000
DEFINITION Sequence 1 from patent US 5952171.
ACCESSION AR073576
VERSION AR073576.1 GI:10000340
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 4951)
AUTHORS McCarthy,S,Anthony, Gearing,D,Paul and Levinson,D,Adam.
TITLE Method for identifying genes encoding secreted or
        membrane-associated proteins
JOURNAL Patent: US 5952171-A 1 14-SEP-1999;
FEATURES
        source
            1..4951
            /organism="unknown"
            location/Qualifiers

BASE COUNT 1152 a 1410 c 1302 g 1087 t

Query Match
Best Local Similarity 100.0%; Pred. No. 1,1e-125;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      230  ccatcgaaggagcgcgcgcctcctcctcagcgcgcgcgcctcctcagtgagcgccatcc 289
          |||
Db      357  CGATTCAGAGGGCTCGCATCTCTCTCAGCGCGCCGCCGCTCACTGAGGGCCATGCC 416
          |||
Qy      290  acgcgcggtgagtcgctctcgcgcgcctccgcgccttgctgctcctcgaactgcgcgc 349
          |||
Db      417  AGCGCGTTGAGTCGCGCTCTGCGCGCTCCGCGCTGCGCTCTGAACTGCTGCC 476
          |||
Qy      350  cgtctagtaagttaaaagctcagatcagacgagcgcccttgcgcgcgccttccttgag 409
          |||
Db      477  CGCTGAGTAAGTTTAAAGCTCAGGTCGAGACCGGGCCCTTTGTCGGCGCTCCCTTGAG 536
          |||
Qy      410  cctactagactcagcgcgcctcctcagcttgcctcagacccttgcctcgaactcga 469
          |||
Db      537  CCACTAGACTCAGCGCGCTCTCAGAGCTTTCCTGACCTGCTGCTCACTACGTA 596
          |||
Qy      470  cttgttctc 478
          |||
Db      597  CTTTGTTTC 605

RESULT 11
AF286077 LOCUS AF286077 8799 bp DNA circular SYN 27-DEC-2000
DEFINITION Expression vector AF286077, complete sequence.
ACCESSION AF286077
VERSION AF286077.1 GI:12034712
KEYWORDS
SOURCE Expression vector AF286077.
ORGANISM Expression vector AF286077.
          artificial sequence: vectors.

REFERENCE 1 (bases 1 to 8799)
AUTHORS Wang,R.F. and Mullins,J.I.
TITLE Mammalian cell/vaccinia virus expression vectors with increased
        stability of retroviral sequences in Escherichia coli: production
        of feline immunodeficiency virus envelope protein
JOURNAL Gene 153 (2), 197-202 (1995)
MEDLINE 95180718
PUBMED 7875588
REFERENCE 2 (bases 1 to 8799)
AUTHORS Mullins,J.I., Hanley,T., Katsack,B.F.C. and Rowen,L.
TITLE Direct Submission
JOURNAL Submitted (11-JUL-2000) Microbiology, University of Washington, HSB
        K-455, Seattle, Washington 98195, USA
        Location/Qualifiers
            1..8799
            /organism="Expression vector AF286077"
            /db_xref="taxon:146704"
            join(1..500,8708..8799)
            /note="vaccinia thymidine kinase gene cluster 5' end"
            46..500
            /gene="thymidine kinase"
            /note="derived from vaccinia; 5' end"
            /pseudo
            502..864
            /note="derived from SV40"
            872..1306
            /note="SR-alpha Hybrid Promoter; formed by HTLV LTR and
            SV40 promoter"
            1344..1892
            /note="derived from EMV"
            1849..1857
            /note="oligopyrimidine tract"
            1917..1934
            /note="polyhistidine tract"
            1960..2007
            /note="similar to tpa 3' end"
            2018..2288
            /note="derived from bovine growth hormone"
            2301..2331
            /note="synthetic"
            2339..2465
            /note="derived from 77"
            2474..2747
            /note="derived from P7.5 vaccinia"
            2777..3143
            /note="murine beta-globin transcriptional regulation unit"
            3189..3851
            /note="DHFR"
            3888..4141
            /note="derived from bovine growth hormone"
            4142..4507
            /note="murine beta-globin transcriptional regulation unit"
            4527..5492
            /gene="neomycin resistance"
            5496..5632
            /note="derived from SV40 hormone"
            5638..5911
            /gene="thymidine kinase"
            /note="derived from vaccinia; 3' end"
            /pseudo
            5638..6265
            /note="vaccinia thymidine kinase gene cluster 3' end"
            6320..6449
            /note="similar to pUC19"
            6453..7047
            /note="p15A intermediate copy number"
            7050..8192
            /note="similar to pUC19"
            complement(7192..8052)
            /codon_start=1
            /product="beta lactamase"
            /protein_id="AAG45964.1"
            /db_xref="GI:12034713"

```

```

/translation="MSIOHFRVALIPFAFACLPVFAHPDETLYKVADEADOLGARVGY
IELDINSKILLESFREPERSFPMSTFKVLGCNAVLSRIDAGEOLGRIRHYSONDIVE
YSPYPERKILTDGNTRELOCSAATMSDNTAANLLITGGPEELAFLEINSGDHVRL
DWEPELNAIPNDERDITPMVAMATTLKRLITGELLTLASROLLIDMEADKVAQPL
LSALPAGWFIADKSGAGERSGIIAALGPDKPSRIIVYITTSQATMDERNROIA
EIGASLIKHM"
misc_feature      8210..8636
                  /note="fl packaging origin"
BASE COUNT      2343 a 2028 c 2168 g 2260 t
ORIGIN
Query Match      9.7%: Score 249; DB 56; Length 8799;
Best Local Similarity 100.0%; Pred. No. 1.2e-125;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 230 cgaatcagagggcgcacatctctctctcctcctcagccgcgcacactgaagccgcacatcc 289
    |||||||
DB 865 CGATTCGAGGGGCTCGCATCTCTCTACAGCGCCGCCGCTCACTGAGCGCCGCTCC 924
QY 290 aagcgcggtgaagtcgcgtctcgcgcctccgcctctgtgctcctcctgaactgcgtccgc 349
    |||||||
DB 925 ACCGCGGTTGAGTCGGCTTCGCCGCCCTCCGCTGTGCTCTCGAAGTCGCTCCGC 984
QY 350 cgcctagtgtaagttaagtcgagtcgagccgcttgcgcggcgctcccttgag 409
    |||||||
DB 985 CGCTAGGTAGTATTAAAGCTCAGGTGAGACCGCGGCTTGTCTCGGCGCTCCCTTGAG 1044
QY 410 cctactaagcagcagcgcgcctcctcagccttgcctgaacctgtcctgaactcact 469
    |||||||
DB 1045 CCACTACGACCTCAGCGCGGCTTCACAGCTTGGCTGACCTGCTTCTCACTACGTACGT 1104
QY 470 cttgtgttc 478
    |||||||
DB 1105 CTTTGTTC 1113

RESULT 12
LOCUS      AB030743      506 bp      DNA      VRL      04-AUG-2000
DEFINITION Human T-cell lymphotropic virus type 1 proviral LTR sequence,
ACCESSION  AB030743
VERSION     AB030743.1 GI:9711249
KEYWORDS   Human T-cell lymphotropic virus type 1 (isolate:Kagoshima HTLV-I)
SOURCE      Human T-cell lymphotropic virus type 1
ORGANISM    Human T-cell lymphotropic virus type 1
REFERENCE   1 (sites)
AUTHORS     Li,H., Fujiyoshi,T., Lou,H., Yashiki,S., Sonoda,S., Cartier,L.,
            Nunez,L., Munoz,I., Horal,S. and Tajima,K.
            The presence of ancient HTLV-I provirus DNA in Andean mummies
            Unpublished (1999)
TITLE       2 (bases 1 to 506)
JOURNAL     Direct Submission
            Submitted (21-JUL-1999) to the DDBJ/EMBL/Genbank databases.
            Hongchuan Li, Kagoshima University, Department of Virology; 8-35-1,
            Sakuragaoka, Kagoshima-city, Kagoshima 890-8520, Japan
            (E-mail:lhcmcd5.kufm.kagoshima-u.ac.jp, Tel:81-99-275-5283,
            Fax:81-99-265-8164)
FEATURES
    source      1..506
                location/Qualifiers
                /organism="Human T-cell lymphotropic virus type 1"
                /proviral
                /isolate="Kagoshima HTLV-I"
                /db_xref="taxon:11908"
                /clone="KAG130"
                /note="seropositive"
LTR          97 a 173 c 121 g 115 t
BASE COUNT
ORIGIN

```

```

Query Match      9.4%: Score 242; DB 58; Length 506;
Best Local Similarity 100.0%; Pred. No. 6.2e-122;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 237 ggggctgcacatctctctcttaagcgccgcgcctcctcctgaagcgccctcaagcgcg 296
    |||||||
DB 230 GGGGCTCGCATCTCTCTCAGCGCGCCGCCCTTACGTAGAGCCGCCATTCACGCGCG 289
QY 297 ttgagtcgctctgcgcctccgcgcctgtgtgagcctcctgaactgcgtccgctcag 356
    |||||||
DB 290 TTGAGTCGCTTTCGCCGCTCCGCCCTGTGTGCTCTCTGAACTGCGTCCGCTTAG 349
QY 357 gtaagttaagcagcagtcgagacggcgccctgttcgcgcctccctgagactaact 416
    |||||||
DB 350 GTAAGTTTAAAGCTCAGGTGAGACCGCGGCTTGTCTCGCGCGCTTGTGAGACTACCT 409
QY 417 agactagccgcgcctcctcaagccttgcctgaacctgttgcctgaactcagcttctgt 476
    |||||||
DB 410 AGACTAGCGCGGCTTCACGCTTGTGCTGACCTGTCTCAACTCTACGTCTTGT 469
QY 477 tc 478
    ||
DB 470 TC 471

RESULT 13
LOCUS      AF014658      508 bp      DNA      VRL      30-AUG-1999
DEFINITION Human T-cell lymphotropic virus type 1 isolate BRAS08, 5' Long
ACCESSION  AF014658
VERSION     AF014658.1 GI:3327436
KEYWORDS   Human T-cell lymphotropic virus type 1
SOURCE      Human T-cell lymphotropic virus type 1
ORGANISM    Human T-cell lymphotropic virus type 1
REFERENCE   1 (bases 1 to 508)
AUTHORS     Yamashita,M., Veronesi,R., Menna-Barreto,M., Harrington,W.J. Jr.,
            Sampaio,C., Brites,C., Badaro,R., Andrade-Filho,A.S., Okhura,S.,
            Igarashi,T., Takeda,J., Miura,T., Chamone,D., Bianchini,O.,
            Jardim,C., Sonoda,S. and Hayami,M.
            Molecular epidemiology of human T-cell leukemia virus type I
            (HTLV-I) Brazil: the predominant HTLV-Is in South America differ
            from HTLV-Is of Japan and Africa, as well as those of Japanese
            immigrants and their relatives in Brazil
            Virology 261 (1), 59-69 (1999)
TITLE       2 (bases 1 to 508)
JOURNAL     Direct Submission
            Submitted (18-JUL-1997) Institute for Virus Research, Kyoto
            University, Shogoin-Kawahara-Machi 53, Sakyo-Ku, Kyoto 606, Japan
FEATURES
    source      1..508
                location/Qualifiers
                /organism="Human T-cell lymphotropic virus type 1"
                /isolate="BRAS08"
                /db_xref="taxon:11908"
LTR          98 a 173 c 123 g 114 t
BASE COUNT
ORIGIN
Query Match      9.4%: Score 242; DB 58; Length 508;
Best Local Similarity 100.0%; Pred. No. 6.2e-122;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 237 ggggctgcacatctctctcctcagcgccgcgcctcctcctgaagcgccctcaagcgcg 296
    |||||||
DB 232 GGGGCTCGCATCTCTTCACGCGCGCCGCCCTTACGTAGAGCCGCCATTCACGCGCG 291
QY 297 ttgagtcgctctgcgcctccgcgcctgtgtgagcctcctgaactgcgtccgctcag 356

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 12, 2001, 12:41:33 ; Search time 167.51 Seconds

(without alignments)
8925.222 Million cell updates/sec

Title: US-09-464-039-8

Perfect score: 2561

Sequence: 1 aggcagaagtgtgcaaacga.....gsgmggraswmwawrrammc 2561

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 678276 seqs, 291890651 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

N.Geneseq_0401:*

- 1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the distribution, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	249	9.7	4951	V27206	Vector pTRAP3 DNA
2	242	9.4	755	N92604	Recombinant DNA ve
3	242	9.4	9045	T42902	DNA sequence which
4	242	9.4	9047	N40080	Sequence complemen
5	241	9.4	633	V35786	Sequence of the sp
6	230	9.0	1675	O50836	HTLV-1 p21X cDNA.
7	230	9.0	1866	O50835	HTLV-1 tax/rex spl
8	188	7.3	633	T97159	SR alpha enhancer/
9	182	7.1	720	N90825	Human T lymphotro
10	182	7.1	720	O68398	HTLV-1 LTR genomic
11	182	7.1	720	T47852	HTLV-I long termin

ALIGNMENTS

12	179	7.0	198	21	C08959	Human secreted pro
13	163	6.4	3796	21	A27831	Vector plasmid pCM
14	163	6.4	6253	20	X08454	AAV vector sequenc
15	163	6.4	6280	20	X08455	AAV vector sequenc
16	163	6.4	6280	20	X08456	AAV vector sequenc
17	163	6.4	6295	19	V02043	Plasmid pMRG3196 e
18	163	6.4	6981	21	Z45931	Nucleotide sequenc
19	163	6.4	7287	19	V02042	Plasmid pMRG3169 e
20	163	6.4	7380	20	X84028	MP9 promoter beta
21	163	6.4	7612	21	Z39629	DNA sequence of pl
22	163	6.4	7852	17	T27556	Shuttle vector pad
23	163	6.4	7892	13	O30906	psAB132. Synthet
24	163	6.4	7897	17	T27555	Shuttle vector pad
25	163	6.4	8236	18	T47202	Recombinant adenov
26	163	6.4	8509	18	T59271	Plasmid pAY.CMVlac
27	163	6.4	8509	20	X33862	AV.CMVlac2 cis pla
28	163	6.4	9077	19	V09005	Vector containing
29	163	6.4	9077	19	V09006	Vector containing
30	163	6.4	9077	21	Z45251	Expression vector
31	163	6.4	9077	21	Z45253	Expression vector
32	163	6.4	10398	17	T15286	Ad.AV.CMVlac2 hybr
33	163	6.4	10398	19	V22130	cDNA sequence of p
34	163	6.4	10398	19	V22129	cDNA sequence of p
35	163	6.4	10398	19	V22139	cDNA sequence of p
36	163	6.4	12135	18	T60555	Plasmid pAdCMVgag-
37	163	6.4	19307	17	T27558	Shuttle vector pad
38	163	6.4	36538	18	T60558	Recombinant trans
39	160	6.2	5068	20	X84027	MP9 promoter GFP
40	127	5.0	304	20	V86402	EST clone AR253.
41	104	4.1	160	21	Z98740	Promoter region fr
42	104	4.1	471	21	Z98741	Promoter region fr
43	103	4.0	4249	19	V63466	Plasmid pCMVIE. C
44	98	3.8	2640	13	O26664	bDAR. Bos taurus.
45	92	3.6	1028	16	T62780	DDP drug resistant

RESULT 1

RESULT	1
V27206	V27206 standard; cDNA; 4951 BP.
ID	V27206;
XX	
AC	V27206;
XX	
DT	12-OCT-1998 (first entry)
XX	
DE	Vector pTRAP3 DNA sequence.
XX	
KW	Secreted protein; membrane-associated protein; protein secretion;
KW	Signal peptide; alkaline phosphatase; pTRAP3; vector; reporter; ss.
XX	
OS	Homo sapiens.
XX	
FT	Key
FT	mat_peptide
FT	Location/Qualifiers
FT	/*tag= a
FT	/product= human placental alkaline phosphatase
PN	WO9822491-A1.
XX	
PD	28-MAY-1998.
XX	
PF	06-NOV-1997; 97WO-US20201.
XX	
PR	19-NOV-1996; 96US-0752307.
XX	
PA	(MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX	
PI	Gearing DP, Levinson DA, McCarthy SA;
XX	
DR	WPI: 1998-312407/27.
DR	P-PDB: W55047.

|||||
Db 352 ggggtcgcattctctctcaacgcgcgcgcctacactgagcgacatccacgcgcg 411
Oy 297 ttgagtcgcgttcctcgcgcctccgcctcttgctctcctgaacgcgcgcgcctcag 356
Db 412 ttgagtcgcgttcctcgcgcctccgcctcttgctctcctgaacgcgcgcgcctcag 471
Oy 357 gtaagttaaaagctcaggtcagacgcgcgcctcttgccgcgcctcccttgagacct 416
Db 472 gtaagttaaaagctcaggtcagacgcgcgcctcttgccgcgcctcccttgagacct 531
Oy 417 agactcagcgcgcctcctcagcgtcttgctcgcgcctcttgctcgaactcagcttctgt 476
Db 532 agactcagcgcgcctcctcagcgtcttgctcgcgcctcttgctcgaactcagcttctgt 591
Oy 477 tc 478
Db 592 tc 593

RESULT 5
V35786
ID V35786 standard; DNA: 633 BP.
XX
AC V35786;
XX
DT 22-SEP-1998 (first entry)
XX
DE Sequence of the specification.
XX
KW Induce; apoptosis; guanine phosphoribosyltransferase; gpt; luciferase;
KW cytokine; tumour necrosis factor; interleukin-1; inhibitory effect;
KW intracellular signal transmission; ss.
XX
OS Unidentified.
XX
PN M09822578-A1.
XX
PD 28-MAY-1998.
XX
PE 12-NOV-1997; 97WO-JP04126.
XX
PR 15-NOV-1996; 96JP-0305163.
XX
PA (CYTO-) INST CYTOSIGNAL RES INC.
XX
PI Nagasawa Y, Yoshida H;
XX
DR WPI: 1998-312464/27.
XX
PT Test system for detecting intra-cellular signal transmission
PT inhibition - using vector containing apoptosis-inhibiting or
PT reporter gene and promoter sequence, used for, e.g. screening for
PT potential anti-inflammatory agents
XX
PS Disclosure: Fig 7; 62pp; Japanese.
XX
CC The present sequence is used in the plasmid vectors of the
CC invention. The vectors which comprise a gene which can induce
CC apoptosis under specific conditions, e.g. guanine
CC phosphoribosyltransferase (gpt), or a reporter gene, e.g. luciferase,
CC where the gene is situated downstream of a promoter which responds to
CC specific extracellular stimulation such as the presence of a cytokine,
CC e.g. tumour necrosis factor (TNF) or interleukin-1. The vector may be
CC used to transform a suitable cell line, such as a cell line which does
CC not produce hypoxanthine-guanine phosphoribosyl transferase (HGPRT). The
CC transformed cells are used to test the inhibitory effect of a gene or
CC substance on intracellular signal transmission.
XX
SQ Sequence 633 BP; 114 A; 210 C; 156 G; 153 T; 0 other;

Query Match 9.4%; Score 241; DB 19; Length 633;

Best Local Similarity 100.0%; Pred. No. 7.3e-102;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 238 ggggtcgcattctctctcaacgcgcgcgcctacactgagcgacatccacgcgcgt 297
Db 355 ggggtcgcattctctctcaacgcgcgcgcctacactgagcgacatccacgcgcgt 414
Oy 298 tgaagtcgcgttcctcgcgcctccgcctcttgctctcctgaacgcgcgcgcctcag 357
Db 415 tgaagtcgcgttcctcgcgcctccgcctcttgctctcctgaacgcgcgcgcctcag 474
Oy 358 taagttaaaagctcaggtcagacgcgcgcctcttgccgcgcctcccttgagacct 417
Db 475 taagttaaaagctcaggtcagacgcgcgcctcttgccgcgcctcccttgagacct 534
Oy 418 gactcagcgcgcctcctcagcgtcttgctcgcgcctcttgctcgaactcagcttctgt 477
Db 535 gactcagcgcgcctcctcagcgtcttgctcgcgcctcttgctcgaactcagcttctgt 594
Oy 478 c 478
Db 595 c 595

RESULT 6
Q50836
ID Q50836 standard; cDNA: 1675 BP.
XX
AC Q50836;
XX
DT 09-MAY-1994 (first entry)
XX
DE HTLV-1 p21X cDNA.
XX
DE HTLV-1; human T cell leukemia virus; PCR; polymerase chain reaction;
KW detection; splice; ss.
XX
KW HTLV-1; human T cell leukemia virus; PCR; polymerase chain reaction;
KW detection; splice; ss.
XX
OS Homo sapiens.
XX
PN JP05244999-A.
XX
PD 24-SEP-1993.
XX
PE 18-DEC-1991; 91JP-0354839.
XX
PR 18-DEC-1991; 91JP-0354839.
XX
PA (SHTO) SHIONOGI & CO LTD.
XX
PI (SHTO) SHIONOGI & CO LTD.
XX
DR WPI: 1993-338952/43.
XX
PT Oligo-nucleotide primer - for detecting mRNA of human T cell
PT leukemia virus 1 by polymerase chain reaction
XX
PS Claim 1; Fig 8; 19pp; Japanese.
XX
CC (Q50836) shows cDNA prepared from RNA extracted from a HTLV-1
CC infected cell. This was amplified by PCR and was used to construct
CC p21X cDNA (Q50836). Primers and probes were then manufactured
CC (Q50837-43), these are useful for the detection of HTLV-1 infection.
XX
SQ Sequence 1675 BP; 349 A; 613 C; 336 G; 377 T; 0 other;

Query Match 9.0%; Score 230; DB 14; Length 1675;
Best Local Similarity 100.0%; Pred. No. 8.5e-97;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 237 ggggtcgcattctctctcaacgcgcgcgcctacactgagcgacatccacgcgcgt 296
Db 1446 ggggtcgcattctctctcaacgcgcgcgcctacactgagcgacatccacgcgcgt 1505
Oy 297 ttgagtcgcgttcctcgcgcctccgcctcttgctctcctgaacgcgcgcgcctcag 356

Db 1506 ttgagtcgcttcgagcgcctccgcctggtgctctcctgaacagcgtccgcgctctag 1565
|||||
Qy 357 gtaagtttaagagctcaggtcagaccggagcctttgtccgagcgtccctctgagactact 416
|||||
Db 1566 gtaagtttaagagctcaggtcagaccggagcctttgtccgagcgtccctctgagactact 1625
|||||
Qy 417 agactaacgcgagctctcagagcgtttgctgaacctgtgtctcaactcta 466
|||||
Db 1626 agactaacgcgagctctcagagcgtttgctgaacctgtgtctcaactcta 1675
|||||

RESULT 7

050835
ID Q50835 standard; cDNA; 1866 BP.
AC Q50835;
XX
XX 09-MAY-1994 (first entry)
DT HTLV-1 tax/rex splice region.
XX
XX HTLV-1; human T cell leukemia virus; PCR; polymerase chain reaction;
KW detection; splice; ss.
XX
XX Homo sapiens.
OS
XX JP05244999-A.
PN
XX 24-SEP-1993.
PD
XX 18-DEC-1991; 91JP-0354839.
PE
XX 18-DEC-1991; 91JP-0354839.
PR
XX 18-DEC-1991; 91JP-0354839.
PT
XX (SH10) SHIONOGI & CO LTD.
PA
XX
XX WPI; 1993-338952/43.
DR
XX
XX Oligo-nucleotide primer - for detecting mRNA of human T cell
PT leukemia virus 1 by polymerase chain reaction
XX
XX
PS Disclosure; Fig 7; 19pp; Japanese.
CC (Q50835) shows cDNA prepared from RNA extracted from a HTLV-1
CC infected cell. This was amplified by PCR and was used to construct
CC p21X cDNA (Q50836). Primers and probes were then manufactured
CC (Q50837-43), these are useful for the detection of HTLV-1 infection.
XX
XX
SQ Sequence 1866 BP; 394 A; 678 C; 386 G; 408 T; 0 other;

Query Match 9.0%; Score 230; DB 14; Length 1866;
Best Local Similarity 100.0%; Pred. No. 8.5e-97;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 237 ggggctgcacatctctctcctcagcgcgcgcgcctactcagagcgccacacgcgcg 296
|||||
Db 1637 ggggctgcacatctctcctcagcgcgcgcgcctactcagagcgccacacgcgcg 1696
|||||
Qy 297 ttgagtcgcttcgagcgcctccgcgcgtgtgtctccttaacagcgtccgcgcttag 356
|||||
Db 1697 ttgagtcgcttcgagcgcctccgcgcgtgtgtctccttaacagcgtccgcgcttag 1756
|||||
Qy 357 gtaagtttaagagctcaggtcagaccggagcctttgtccgagcgtccctctgagactact 416
|||||
Db 1757 gtaagtttaagagctcaggtcagaccggagcctttgtccgagcgtccctctgagactact 1816
|||||
Qy 417 agactaacgcgagctctcagagcgtttgctgaacctgtgtctcaactcta 466
|||||
Db 1817 agactaacgcgagctctcagagcgtttgctgaacctgtgtctcaactcta 1866
|||||

RESULT 8

T97159
ID T97159 standard; DNA; 633 BP.
XX
XX T97159;
AC
XX
XX 11-MAY-1998 (first entry)
DT
XX
XX SR alpha enhancer/promoter.
DE
XX
XX Vaccine; B-cell malignancy; lymphoma; leukemia; tumour;
KW gene amplification; immunotherapy; therapy; SV40; promoter;
KW enhancer; ds.
XX
XX Chimeric - Rhesus macaque polyoma virus.
OS
XX Chimeric - Human T cell leukemia virus type 1.
XX
XX WO9741244-A1.
PN
XX
XX 06-NOV-1997.
PD
XX
XX 25-APR-1997; 97WO-US07039.
PE
XX
XX 06-DEC-1996; 96US-0761277.
PR
XX 01-MAY-1996; 96US-0644664.
XX
XX (GENI-) GENITOPE CORP.
PA
XX
XX
XX Denney DW;
PI
XX
XX WPI; 1997-549743/50.
DR
XX
XX Multivalent vaccine to treat B cell lymphoma or leukemia -
PT comprises at least 2 different recombinant variable regions of
PT immunoglobulin molecules derived from B cell lymphoma cells
XX
XX
XX Example 1; Page 104; 177pp; English.

CC This DNA sequence comprises the HindIII/XhoI fragment of the
CC SR alpha enhancer/promoter in plasmid pCDL-SR alpha 296. The
CC SR alpha enhancer/promoter is composed of human T cell leukemia
CC virus 1 5' untranslated sequences and the SV40 enhancer. It is
CC reported to increase expression from the SV40 enhancer/promoter by
CC 10-fold in host cells, and is active in a broad range of cell
CC types. The SR alpha enhancer/promoter has been utilised in
CC expression vectors designed for efficient expression of genes in
CC eukaryotic cells. The invention provides vectors and improved
CC methods for the expression and co-amplification of genes encoding
CC recombinant proteins in cultured cells. The methods permit the
CC isolation of cell lines which have co-amplified input recombinant
CC sequences which encode an amplifiable marker, one or more
CC expression vectors encoding a protein of interest and optionally a
CC selectable marker. The amplified cells provide large quantities
CC of recombinant proteins suitable for immunotherapy for treatment of
CC lymphomas and leukaemias. The methods permit the production of
CC custom vaccines, including multivalent vaccines that reflect the
CC degree of somatic variation found in a patient's tumour.
XX
XX
SQ Sequence 633 BP; 114 A; 210 C; 155 G; 154 T; 0 other;

Query Match 7.3%; Score 188; DB 18; Length 633;
Best Local Similarity 99.6%; Pred. No. 2.5e-77;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 240 gctgcacatctcctcctcagcgcgcgcgcctactcagagcgccacacgcgcgtg 299
|||||
Db 359 gctgcacatctcctcctcagcgcgcgcgcctactcagagcgccacacgcgcgtg 418
|||||
Qy 300 agtcgagcttcgagcgcctccgcgcgtgtgtgtcctccttaactcgtccgcgctcagta 359
|||||
Db 419 agtcgagcttcgagcgcctccgcgcgtgtgtgtcctccttaactcgtccgcgctcagta 478
|||||

Sequence 720 BP; 138 A; 252 C; 157 G; 173 U; 0 other;

Matches 131; Conservative 51; Mismatches 0; Indels 0; Gaps 0;

Db 239 uc 240

ID T47852 standard; RNA; 720 BP.

AC T47852;

DT 21-MAY-1997 (first entry)

DE HTLV-I long terminal repeat region oligonucleotide.

KW HIV; human immunodeficiency virus; antisense; replication inhibitor; infection resistant; retrovirus; immunomodulators; disease; viral load

KM Lymphadenopathic; HTLV-I; human T-cell lymphotropic virus type I; AIDS; acquired immune deficiency syndrome; PCP; polypurpura chain reaction; dc

XX Human T-cell lymphotropic virus type T

AA
PN
US5580761-A.

PD 03-DEC-1996

PF 16-FEB-1988; 88US-0156188.

PR 23-MAR-1994; 94US-0217210.

PR 29-JUL-1991; 91US-0739718.

PA (GREW) GREATBATCH GEN-AID LTD.
XX

PI Greatbatch W, Sanford JC;
VY

DR WPI; 1997-033571/03.
XX

PT transformation in vitro with DNA construct encoding retroviral RNA

PI able to inhibit replication
XX

XX

Sequence 720 BP; 138 A; 252 C; 157 G; 173 U; 0 other;

Matches 131; Conservative 51; Mismatches 0; Indels 0; Gaps 0;

Db 239 uc 240

ID	standard; cDNA; 198 BP
C08959	

AC C08959;

DT 06-OCT-2000 (first entry)
yy

DE	Human secreted protein 5' EST, SEQ ID NO: 13034.
YY	

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation
gene therapy; chromosome mapping; ss

Homo sapiens -
XX
OS

XX
PN
EP1033401-A2

AA PD 06-SEP-2000.

21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PA (GEST) GENSET

PI Dumas Milne Edwards J, Duclert A, Giordano J

DR WPI; 2000-500381/45.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
BT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

PT diagnostic, forensic, gene therapy and chromosome mapping procedures

PS Claim 1; SEQ ID 13034; 71pp + CD-ROM; English
XX

CC The present sequence is one of a large number of 5' ESTs derived from


```

XX Humanised green fluorescent protein - used to measure gene
PT expression and identify transformed cells
XX
XX Claim 2; Page 132-136; 152pp; English.
XX
CC Humanised green fluorescent protein (gfp) genes can be used to
CC identify transformed cells, to measure gene expression in vitro and
CC in vivo, to label specific cells in multicellular organisms (e.g. to
CC study cell lineage/s), to label and locate fusion proteins, and to
CC study intracellular trafficking. Commonly used reporter genes include
CC beta-galactosidase, firefly luciferase, alkaline phosphatase,
CC chloramphenicol acetyltransferase (CAT), and beta glucuronidase
CC (GUS). However, these have limitations in their use. Frequently,
CC these reporter genes require the addition of a substrate and the
CC size of certain proteins means that the expression of reporter
CC fusion proteins can be difficult. The light stimulated GFP
CC fluorescence is species independent and does not require any
CC cofactors substrates or additional gene products from Aequorea
CC victoria an as the GFP genes have been humanised, they are
CC expressed at sufficient levels to be detectable in human cells,
CC unlike previous GFP proteins.
CC
SQ Sequence 6253 BP; 1505 A; 1624 C; 1681 G; 1443 T; 0 other;
XX
XX
Query Match 6.4%; Score 163; DB 20; Length 6253;
Best Local Similarity 100.0%; Pred. No. 8.4e-66;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 513 ctgaaacccagaagtgtaagtttagtctttttgtctttatttcaggtccg 572
DB 804 ctgaaacccagaagtgtaagtttagtctttttgtctttatttcaggtccg 863
OY 573 gatccggtgtgtgcaaatcaagaactgctcctcagtgagttgtccttactctag 632
DB 864 gatccggtgtgtgcaaatcaagaactgctcctcagtgagttgtccttactctag 923
OY 633 gctctacggaagtgttactctgtctctaaagtcgcggaatt 675
DB 924 gctctacggaagtgttactctgtctctaaagtcgcggaatt 966
XX
RESULT 15
X08455
ID X08455 standard; DNA: 6280 BP.
XX
AC X08455;
XX
DT 28-JUN-1999 (first entry)
XX
DE AAV vector sequence comprising humanised green fluorescent protein.
XX
KW Green fluorescent protein; gfp; jellyfish; Aequorea victoria;
KW humanisation; reporter gene; substrate; cofactor; beta galactosidase;
KW firefly luciferase; alkaline phosphatase;
KW chloramphenicol acetyltransferase; CAT; beta glucuronidase; GUS; ss.
XX
OS Synthetic.
XX
XX Key Location/Qualifiers
FT CDS 988..1731
FT /tag= a
FT /product= "Humanised green fluorescent protein"
XX
XX MO9903997-A1.
XX
XX 28-JAN-1999.
XX
XX 16-JUL-1998; 98WO-US14692.
XX
XX 16-JUL-1997; 97US-0893327.
XX

```

```

PA (UYFL ) UNIV FLORIDA.
XX
XX Hauswirth W, Muzyczka N, Zolotukhin S;
XX
XX WPI; 1999-132241/11.
XX
XX P-PSDB; w96329.
XX
XX
XX Claim 3; 137-141; 152pp; English.
XX
XX
CC Humanised green fluorescent protein (gfp) genes can be used to
CC identify transformed cells, to measure gene expression in vitro and
CC in vivo, to label specific cells in multicellular organisms (e.g. to
CC study cell lineage/s), to label and locate fusion proteins, and to
CC study intracellular trafficking. Commonly used reporter genes include
CC beta-galactosidase, firefly luciferase, alkaline phosphatase;
CC chloramphenicol acetyltransferase (CAT), and beta glucuronidase
CC (GUS). However, these have limitations in their use. Frequently,
CC these reporter genes require the addition of a substrate and the
CC size of certain proteins means that the expression of reporter
CC fusion proteins can be difficult. The light stimulated GFP
CC fluorescence is species independent and does not require any
CC cofactors substrates or additional gene products from Aequorea
CC victoria an as the GFP genes have been humanised, they are
CC expressed at sufficient levels to be detectable in human cells,
CC unlike previous GFP proteins.
CC
SQ Sequence 6280 BP; 1515 A; 1627 C; 1692 G; 1446 T; 0 other;
XX
XX
Query Match 6.4%; Score 163; DB 20; Length 6280;
Best Local Similarity 100.0%; Pred. No. 8.4e-66;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 513 ctgaaacccagaagtgtaagtttagtctttttgtctttatttcaggtccg 572
DB 804 ctgaaacccagaagtgtaagtttagtctttttgtctttatttcaggtccg 863
OY 573 gatccggtgtgtgcaaatcaagaactgctcctcagtgagttgtccttactctag 632
DB 864 gatccggtgtgtgcaaatcaagaactgctcctcagtgagttgtccttactctag 923
OY 633 gctctacggaagtgttactctgtctctaaagtcgcggaatt 675
DB 924 gctctacggaagtgttactctgtctctaaagtcgcggaatt 966
XX

```

Search completed: June 12, 2001, 14:19:58
Job time: 5905 sec

into the Not I and Hind III sites of the Lactid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 150 a 64 c 89 g 172 t 7 others

Query Match 12.7%; Score 326; DB 156; Length 482;
Best Local Similarity 100.0%; Pred. No. 4.2e-161;
Matches 326; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2137 ttataagatagcagcgttctgtcggaaagatagaaattctcttaagaacttgaa 2196
Db 1 ttatagagatagcagcgttctgtcggaaagatagaaattctcttaagaacttgaa 60
Oy 2197 ttgtaattaaatggcaagcctaaacataagctcttaagtgagattctaaagacg 2256
Db 61 ttgtaattaaatggcaagcctaaacataagctcttaagtgagattctaaagacg 120
Oy 2257 tctgtgtttatattcaagggttaacccttgagcctcatctcatctgctt 2316
Db 121 tctgtgtttatattcaagggttaacccttgagcctcatctcatctgctt 180
Oy 2317 tctcaagaagaatgtttggcgagcagctcagatcagcagcagtaaaatagctcttcaa 2376
Db 181 tctcaagaagaatgtttggcgagcagctcagatcagcagcagtaaaatagctcttcaa 240
Oy 2377 atctctgtcagatgaatgaagtagctgttcttaaatcttaagtttgattgattat 2436
Db 241 atctctgtcagatgaatgaagtagctgttcttaaatcttaagtttgattgattat 300
Oy 2437 actaatgaagaatcttaagtattt 2462
Db 301 actaatgaagaatcttaagtattt 326

RESULT 14
LOCUS AA622988/c 386 bp mRNA EST 21-OCT-1997
DEFINITION np58h04.s1 NCI-CGAP Br2 Homo sapiens CDNA clone IMAGE:1130551 3'
sequence similar to WP:CI7G10.8 CE02490 ALCOHOL DEHYDROGENASE ;, mRNA

ACCESSION AA622988
VERSION AA622988.1 GI:2526864
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL COMMENT
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.

REFERENCE
AUTHORS CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LMW at:
www-bio.llnl.gov/dbip/image/image.html
Insert Length: 937 Std Error: 0.00
Seq primer: -40ml3 fwd. ER from Amersham
High quality sequence stop: 342.
Location/Qualifiers
1. 386

FEATURES
SOURCE
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1130551"

/clone.lib="NCI CGAP Br2"
/sex="female, pooled"
/tissue_type="breast"
/lab_host="DH10B"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from pooled bulk
breast tumor tissue, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified p773
vector. This library is the normalized version of
NCI CGAP Br1.1. Library was constructed by Bento Soares
and M. Fatima Bonaldo."

BASE COUNT 99 a 92 c 58 g 137 t
ORIGIN

Query Match 12.7%; Score 325; DB 9; Length 386;
Best Local Similarity 100.0%; Pred. No. 1.4e-160;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1694 agaaacatttagaattgttaaggaactctcagatgattgtttaagcacttaagc 1753
Db 386 agaaacatttagaattgttaaggaactctcagatgattgtttaagcacttaagc 327
Oy 1754 aatctatcgtttgacctccggtgaagatgttgcaagtgattcttgatcgaag 1813
Db 326 aatctatcgtttgacctccggtgaagatgttgcaagtgattcttgatcgaag 267
Oy 1814 caagggttggaatgtcagatagcagcctctcagcagcagatgttgatagat 1873
Db 266 caagggttggaatgtcagatagcagcctctcagcagcagatgttgatagat 207
Oy 1874 gactactgtagcttctgtaaaatgtttcagggaaactaaacacaaatggcattcat 1933
Db 206 gactactgtagcttctgtaaaatgtttcagggaaactaaacacaaatggcattcat 147
Oy 1934 gtcaaggaaattgaagatgaaggaacatggccctagaaacaaattggaagactat 1993
Db 146 gtcaaggaaattgaagatgaaggaacatggccctagaaacaaattggaagactat 87
Oy 1994 gaacagatgaatgcagactgtga 2018
Db 86 gaacagatgaatgcagactgtga 62

RESULT 15
LOCUS 219446 340 bp mRNA EST 10-FEB-1993
DEFINITION HSB28F112 STRATAGENE Human skeletal muscle CDNA library, cat.
#936215. Homo sapiens CDNA clone 28F11, mRNA sequence.

ACCESSION 219446
VERSION 219446.1 GI:29287
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

REFERENCE
AUTHORS CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LMW at:
www-bio.llnl.gov/dbip/image/image.html
Insert Length: 937 Std Error: 0.00
Seq primer: -40ml3 fwd. ER from Amersham
High quality sequence stop: 342.
Location/Qualifiers
1. 386

JOURNAL COMMENT
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.

FEATURES
SOURCE
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1130551"

THIS PAGE BLANK (USPTO)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☒ **SKEWED/SLANTED IMAGES**
- ☒ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)